UNDER PRESSURE: TEACHER EXPECTATIONS AND STUDENT ACHIEVEMENT IN THE ERA OF SCHOOL ACCOUNTABILITY

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Jennifer J. Todd

May 2011

UNDER PRESSURE: TEACHER EXPECTATIONS AND STUDENT ACHIEVEMENT IN THE ERA OF SCHOOL ACCOUNTABILITY

Jennifer J. Todd, Ph. D.

Cornell University 2011

This dissertation seeks to answer the question: Does pressuring teachers to raise their expectations increase student achievement? Drawing from the sociology of education, educational psychology, and the research on accountability systems, the author constructs a more comprehensive model of the association between teacher expectations, accountability interventions, and student achievement than has been offered in prior research. The author argues that prior research on accountability interventions focuses on the direct relationship between accountability testing and student achievement, but ignores how teachers mediate this association. To this end, prior research ignores the important role that teachers may play in communicating expectations shaped by accountability policies.

Using data drawn from the Education Longitudinal Study matched to a unique state-level accountability dataset, this dissertation offers a systematic assessment of how public school teachers respond to accountability interventions with regard to their expectations for students. Findings show that teacher expectations are important predictors of student achievement, which is consistent with prior research. What prior scholars and policymakers have failed to appreciate, however, is that pressuring teachers to raise their expectations has unanticipated and counterproductive consequences on the very students the policies are intended to help. Rather than raising their expectations of students, teachers appear to use the information gathered

from tests to lower their expectations or even to justify their already low expectations of students, especially for low-performing students. In spite of their lower expectations, the association between teacher expectations and student achievement is stronger in these states because teachers adjust their expectations of students. Further analysis shows that a student's race is an important determinant of teacher expectations. Teachers hold black and Hispanic students to lower standards than their white and Asian peers, and these low expectations contribute to the achievement gaps between white, black, and Hispanic students.

BIOGRAPHICAL SKETCH

Jennifer J. Todd was born in Wooster, OH in February, 1981. She earned a Bachelor of Arts in Sociology from Cornell University in 2003.

To Breean, for being so absolutely fabulous.

ACKNOWLEDGMENTS

I would like to thank all those who have assisted me while completing my dissertation. I am forever indebted to my advisor, Steve Morgan, for his guidance and support. I would also like to thank my special committee members, Kim Weeden, Shelley Correll, and David Strang for their support as well as, and for making graduate school a time not to be forgotten.

To thank my fellow Cornellians—Alexa Yesukevich, Cate Taylor, Christin Munsch, Dafna Gelbgiser, Emily Hoagland, In Paik, Jessica Houston-Su, Jung Mee Park, Laura Ford, Sarah Thébaud, Theo Leenman, Trey Spiller, and Youngjoo Cha—without you I would be lost. I cannot thank you enough for your helpful comments throughout the writing process or for your friendship. I would also like to thank Kelly Andronicos for her friendship, advice, and the many laughs we shared in the Center for the Study of Inequality.

I would also like to thank my family for your support and for helping me get through the latter stages of writing. I would like especially like to thank Breean Walas for all her sage advice, and for reminding me that anything is possible.

Finally, I would like to thank all the teachers I have had over the years who had high expectations of me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

This dissertation seeks to answer the question: Does pressuring teachers to raise their expectations increase student achievement? I draw from the sociology of education, educational psychology, and the research on accountability systems to construct a model that identifies the ways in which accountability interventions affect teacher expectations and student achievement. In the remainder of this chapter, I provide a theoretical explanation for the importance of teacher expectations and why the narrow focus of the extant research on accountability systems leaves a gap in the literature.

TEACHER EXPECTATIONS, STUDENT ACHIEVEMENT, AND THE INCREASING RELIANCE ON ACCOUNTABILITY TESTING

Since the 1940s, sociologists and education scholars have been interested in teacher expectations as determinants of student achievement. Social psychological and educational psychological research has been particularly interested in Merton's (1948) concept of a "self-fulfilling prophecy," which he defines as, "a *false* definition of the situation evoking a new behavior which makes the originally false conception come *true*" (193). This stream of research investigates questions such as: Does raising teacher expectations increase student achievement? And if so, does increasing teacher expectations affect students differently? At the same time, sociological research has also examined the student-teacher relationship extensively. Together, this research offers insight into the dynamics of how teacher expectations are formed and communicated to students, as well as how students respond. Findings overwhelmingly suggest that teacher influence matters, but that teachers hold students to different standards and that these differential expectations may not be related to student effort and ability (Ainsworth-Darnell and Downey 1998; Downey and Pribesh 2004; Ferguson 2000; Ferguson 1998; Porter, Floden, Freeman, Schmidt, and Schwille 1988; Prime and Miranda 2006). While the evidence suggests consistently that teachers hold students to different standards, the extent to which teachers' low expectations contribute to the achievement gap remains unknown after decades of research.

With the enactment of No Child Left Behind (NCLB) in 2002, studying these relationships is more important than ever. NCLB uses a high-stakes accountability system to monitor student achievement and to define achievement expectations. The assumption is that, for many low-performing schools, the mere definition of achievement expectations should pressure teachers to raise their own expectations of students and put an end to what then-President George W. Bush called the "soft bigotry of low expectations."

NCLB's emphasis on accountability is not new to education policy. Since the 1970s, all states have taken steps to raise student achievement by defining content standards and implementing accountability testing. For example, starting in 1977, Florida required that students pass a minimum competency test to graduate from high school. Florida has since revised its content standards and increased accountability standards several times and has also implemented a voucher program that gives students in low-performing schools the option to enroll in a better school. Texas, another early adopter, used tests to monitor student progress and required students to pass an exit exam in order to graduate. Texas was one of the first states to use test scores to sanction poorly performing schools or reward schools that exceeded achievement standards.

Kentucky, by contrast, joined the accountability systems movement much later, but it did so with rigor when it implemented the 1990 Kentucky Education Reform Act (KERA). Through KERA, Kentucky implemented an accountability system that used test scores to monitor student performance on clear state-defined content standards and to target teachers, sanctioning those whose students performed poorly or otherwise rewarding those whose students surpassed achievement expectations. At the same time, some states used tests for diagnostic purposes only (e.g., Iowa) or otherwise deferred to districts to define their own accountability standards (e.g., Nebraska). Over the years, states have revised their achievement expectations and testing requirements in response to their own students' needs as well as evidence gathered from other states, yielding a diverse landscape of accountability systems. One important feature of NCLB, then, is that it mandated state-based legislation on accountability systems, variability that this project will leverage.

Evidence of the impact of state-based accountability interventions such as these is inconclusive (see Hamilton 2003 for an extensive review). Some scholars claim that accountability interventions raise student test scores (Carnoy and Loeb 2002; Roderick, Jacob, and Bryk 2002). Others point out that these gains may be driven by increased retention and drop out rates (Darling-Hammond 2003; Warren, Jenkins, and Kulick 2006), as well as increased rates of students enrolled in special education, where they are likely to be exempt from testing requirements (Amrein-Beardsley and Berliner 2003; Figlio and Getzler 2006).

In this dissertation, I argue that the research on accountability interventions fails to understand the association between accountability interventions and student achievement because it ignores the importance of teachers. The success of accountability interventions is dependent, in part, upon teachers adjusting their expectations in response to state provisions and then successfully communicating them to students. In suggesting that expectations should be uniform and high, accountability systems implicitly discourage teachers from adjusting expectations to the specific features of students, real or inferred. Most research on accountability systems focuses on the direct relationship between accountability testing and student achievement, but ignores how teachers mediate this association. Researchers who examine the impact of accountability interventions on student achievement thus impose one of two assumptions: (1) that students respond directly to accountability expectations and teachers are not important mediators; or (2) that the findings are indirect evidence of teacher effectiveness. In either case, such studies ignore the important role that teachers may play in communicating expectations shaped by accountability policies.

My dissertation seeks to fill this gap in the literature while also addressing the literature on teacher expectation effects: Does pressuring teachers to raise their expectations increase student achievement? Using data drawn from the Education Longitudinal Study matched to a unique state-level accountability dataset that I constructed, I use advanced quantitative methods to examine (1) the association between teacher expectations and student achievement, (2) how accountability interventions impact the student-teacher relationship, and (3) how accountability systems do or do not contribute to eliminating the race achievement gap.

OVERVIEW OF SUBSEQUENT CHAPTERS

Chapter 2 introduces the data used to answer my research questions. I argue that prior research on accountability interventions is limited because it employs data that focus specifically on the association between accountability testing and student outcomes. For this dissertation, I draw the primary data from the 2002, 2004, and 2006 waves of the Education Longitudinal Study (ELS). The ELS is a large, nationally representative dataset that follows a cohort of tenth grade students enrolled in public and private schools during the 2001-02 academic school year. The 2002 base-year wave includes seven components: questionnaires completed by the student, parent, math teacher, English teacher, school administrator, and librarian; and included a school facilities checklist. Together, these questionnaire components provide rich information about students and their teachers, schools and communities. The data also include results for high-quality achievement tests, enabling analysis of the effects of teacher expectations on student achievement.

The ELS data are matched to a unique state-level accountability systems dataset that I compiled using information from each state department of education and other sources. The dataset that I constructed focuses not on the impact of sanctions and rewards on student achievement, as does much of the extant research, but on the type of information states collect in their accountability systems. This allows me to explore how information type affects teacher expectations of students and the effects of those expectations on student achievement. These data, when linked to ELS data, allow me to evaluate heretofore untested hypotheses and assumptions about the relationship between accountability standards and student outcomes, as well as how teachers respond and adjust to expectations standards.

In *Chapter 3*, I address a classic question in the sociology of education and in the current policy environment: Do teachers influence achievement by communicating expectations to students? Increasing evidence suggests that teacher expectations of students' future educational attainment matter, but the association is affected by the context in which students and teachers interact. I construct a more comprehensive model of the association between math teacher expectations and student achievement in math than has been offered in prior research. [Consistent with other research, I analyze math because, unlike reading, standardized tests are a more reliable measure of student achievement.]

I first identify the determinants of teacher expectations of students' future educational attainment, and then examine the association between teacher expectations and student performance. I show that although student characteristics are the most important determinants of teacher expectations, teacher characteristics and school composition mediate how teachers perceive students. I then show that teacher expectations are strongly associated with math test scores and enrollment status in 2004 (i.e., graduate on time, held back a grade, or dropped out), net of the determinants of teacher expectations.

Chapter 4 examines how teachers respond to state-level pressure to raise their expectations. To date, research focuses primarily on the direct impact of accountability interventions on student achievement. There is some research that examines how teachers respond to accountability interventions, but not how teachers then communicate accountability expectations to students. Findings from this research suggest that teachers tend to strongly oppose the use of tests as a means of measuring student achievement because state-level standards interfere with their teaching agendas. However, although tests are intended for use as direct measures of student achievement, they are also indicators of teacher effectiveness, and therefore may be threatening to teachers. Findings from research on accountability systems indicate that the more pressure teachers feel to meet expectations, the more likely they are to compromise or adjust their teaching plans in order to prepare students for the tests (Stecher, Barron, Chun, and Ross 2000).

Even if teachers do not believe that tests are necessary to monitor student attainment, tests can still considered to be used to help teachers identify students in need of additional help (Shepard and Dougherty 1991). Furthermore, attaching sanctions to test scores introduces an incentive for teachers and students to want to perform well. Accordingly, whereas prior research focuses primarily on the impact of sanctions and rewards on student achievement, I focus more attention on how the type of information gathered affects teachers' influence on students. I address three questions: (1) How do teachers respond to accountability interventions? (2) Do teachers communicate these expectations to students? and (3) What are the consequences for student learning for low- and high-performing students?

I show that teachers respond adversely to accountability expectations, such that stronger accountability systems lead teachers to decrease, rather than increase, their expectations of students. I also show that there is little evidence of an accountability effect on student achievement or teachers effectively communicating accountability expectations to students. I posit that teachers selectively adjust their expectations, such that they more strongly predict student achievement for low-performing students than for high-performing students. I show that, in states with strong accountability interventions, the lowest performing students are subject to particularly low teacher expectations. In other words, teachers appear to calibrate their expectations of low-performing students so that they are more realistic, or less hopeful. To this end, I maintain that low-performing students who adopt their teachers' low expectations as their own suffer from the "harsh reality of diminished expectations" rather than the "soft bigotry of low expectations."

Chapter 5 speaks to education policymakers' concern that teachers hold black and Hispanic students to lower standards than they do white students. A primary goal of accountability systems is to pressure teachers to raise their expectations of traditionally low-performing groups, especially black and Hispanic students. Indeed these accountability systems are often justified with reference to the black-white test score gap, one of the most extensively studied group differences in education research and a perennial concern of education policymakers. The Hispanic-white gap is now attracting a growing amount of scholarly interest. Overall, findings from this research suggest that teachers tend to hold black and Hispanic students to lower standards than they do their white and Asian peers, which some argue contributes to the achievement gap.

In this chapter, I argue that student race does in fact determine teacher expectations and that the ways in which teachers mediate accountability expectations differs by student race. I show that teacher expectations are lowest for black and Hispanic students, even after taking into account prior academic performance. I also show that student race affects how teachers respond to accountability testing information. I argue that the findings presented in this chapter suggest that student race should not be overlooked when considering how teachers communicate accountability expectations to students.

Chapter 6 summarizes the findings from the three empirical chapters and discusses the policy implications of increasing reliance on accountability testing.

CHAPTER 2

DATA AND METHODS

The primary objectives of this dissertation are to model the association between teacher expectations and student achievement (e.g., math test scores, as well as the likelihood being held back a grade, dropping out of school, or graduating on time), and to investigate whether teachers respond to state-level pressure to raise their expectations of students. This chapter describes the sources of data used to answer my research questions.

DATA

Education Longitudinal Study

The student-level data were drawn from the 2002, 2004, and 2006 waves of the Education Longitudinal Study of 2002 (ELS), which is sponsored by the National Center for Educational Statistics of the U.S. Department of Education. The ELS uses a two-stage sampling method that first randomly selects public and private high schools and then randomly selects a within-school sample of high school sophomores. The result is a nationally representative sample of 15,244 tenth grade students enrolled in 751 public and private high schools during the 2001-02 academic year.[1] The base-year respondents were then followed and resurveyed in the spring of 2004 and again in 2006, with future follow-up waves scheduled.

For the 2002 base-year wave, questionnaires were distributed to students as well as a parent, math teacher, English teacher, school administrator, librarian, and a school facilities checklist. Collectively, the questionnaires provide rich information

[1] Two respondents who participated in the 2002 base-year wave sample were subsequently removed from the sample for the follow-up data.

about the students, their schools and the communities in which they live, thus making the ELS an ideal source of data to investigate the student-teacher relationship.

The 2004 wave (or first follow-up) was conducted during the spring semester of the 2003-04 academic school year.[2] Whereas the base-year wave gathered information from seven different sources, the first follow-up wave gathered information primarily from students and school administrators. All students completed a student questionnaire similar to the base-year wave student questionnaire. Students who were a part of the freshened sample and base-year wave respondents whose enrollment status changed between the 2002 and 2004 waves (i.e., they dropped out, graduated early, were homeschooled, or transferred schools) were administered a specialized questionnaire component specific to their 2004 enrollment status.

For the 2006 wave (or second follow-up), most respondents had graduated from high school and were enrolled in college, but some were still enrolled in high school or were in the labor market. The 2006 questionnaire included four subsections (i.e., high school, postsecondary education, employment, and community), which makes it possible to track the respondents to their respective destinations over time.

State-Level Accountability Interventions

Whereas the ELS data provide rich information about students and their teachers, the accountability testing information is limited and sparse. For the base-year wave, school administrators were asked a variety of questions about

[2] In addition to the base-year wave participants, the first follow-up wave includes an additional random sample of 953 twelfth grade students drawn from within the base-year wave schools. The additional sample of students includes students who (1) were eligible and did not participate in the base-year wave (N = 649), (2) were ineligible during the base-year wave but became eligible during the 2004 wave (N = 126), or (3) were enrolled in the twelfth grade in the spring of 2004 (N = 178). The addition of the freshened sample makes the 2004 wave nationally representative of students enrolled in the twelfth grade in 2004. Accordingly, the ELS may be used for either longitudinal or cross-sectional purposes. The sample of freshened students is not included in my analysis sample because it does not have teacher information, which was gathered during the base-year wave.

accountability testing in their schools.[3] I do not use this information because there is too much missing information and variation across schools and between states to construct a complete dataset of 2001-02 accountability system requirements for the schools that participated in the base-year wave. Therefore, using information I gathered from each state department of education as well as several additional sources, I constructed a new dataset that captures a variety of the 2001-02 state accountability system characteristics for all 50 states and the District of Columbia. [See Chapter 4 for a more detailed discussion of the testing movement and issues associated with modeling the impact of accountability interventions on teacher expectations and student achievement.]

Implementation. The first variable identifies the states that were early participants in the movement to use tests to monitor student achievement, either for assessment or accountability purposes. Tests have been used for over a century, but it was not until the 1980s that they started to gain legitimacy as a means of holding students accountable to state-defined achievement expectations (Hamilton 2003; Linn 2000). By the late 1990s, most states had implemented an accountability system, and the few that did not have since followed suit in accord with NCLB requirements. Using information drawn from the Office of Technology Assessment (1992), I constructed a variable that identifies the states that were early to implement an accountability system as opposed to those that were late to implement, where early

[3] The base-year wave questionnaire asked school administrators a series of questions pertaining to the use of competency tests. In particular, they were asked if competency tests were administered for grades 8 though 12 the grade levels. If yes, administrators were then asked to identify which subjects were tested for each grade—i.e., math, English, science, and history/social studies (BYA33AA-BYA33FE), and if the competency tests were state, district, or school requirements (BYA34). Administrators were also asked if the competency tests were aligned to content standards or curriculum (BYA35), the percentage of students who failed the test (i.e., scores were not proficient) in any or all of the subject areas on their first attempt (BYA36). Finally, administrators were asked whether students who failed any test were permitted to retake the test and/or were offered educational support programs to help them pass the test (BYA37A-BYA37F).

implementers are defined as having an accountability system prior to 1992.[4] Evidence suggests that test scores increase upon the introduction of accountability expectations, but the effect decreases over time as accountability expectations gain legitimacy and the fear of sanctions decreases (Hanushek and Raymond 2003). Both students and teachers make initial adjustments to newly defined expectations, but once the expectations become more legitimized or realistic, accountability systems lose their impact.

Testing Characteristics. Before states administer assessments, they must first decide what it is that they want to assess or measure. States determine the type of tests they use, what grades and subjects to test students, and whether to attach sanctions to performance. I gathered 2001-02 state and school report cards from each state department of education website, including the District of Columbia. If report card data were not readily available, I requested them via email contact or phone. For each grade, I extracted information about the subject tested, the type of test administered—i.e., norm-referenced tests (NRT), criterion-referenced tests (CRT), or performance assessment tests (PAT)—and whether the test was a required exit exam. Although I gathered information for grades k-12, I restricted the data for this project to testing requirements for grades 9 to 12 because I am most interested in the impact of state-level testing on teacher expectations from the ELS.[5]

I then constructed a series of key variables that capture state-defined testing requirements for k-12 for the 2001-02 academic school year. Four dichotomous

[4] The OTA (1992) is one of the earliest reports assessing the adoption of state accountability systems and is frequently cited in the accountability systems literature. For a more descriptive review of the adoption of assessment/accountability systems, see Linn (2000).
[5] I tested the accuracy of the accountability variables by comparing the information I gathered to information gathered by the Council of Chief State School Officers (2002) and Carnoy and Loeb (2002), both of which offer similar information for the 2000-01 academic school year, and exit exam information was compared to testing information available at the Minnesota Population Center website (http://www.hsee.umn.edu/).

variables identify the types of tests that states used between grades 9 and 12: (1) *NRTs*, (2) *CRTs*, (3) *PATs*, and/or (4) *exit exams* at any point between grades 9 and 12.[6] I use these variables to assess how testing rigor affects teachers and students. Four additional variables capture the number of years students are tested in mathematics, English, science, and history/social studies, ranging from 0 to 4 years. From these four variables, I constructed an index of the *frequency of testing* within each state. The resultant index has a Cronbach's α of .87, which is quite strong under statistical terms (Nunnally and Bernstein 1994). I then matched the state-level accountability data that I constructed with the ELS data, creating a unique dataset that allows me to model the association between accountability interventions, teacher expectations, and student achievement.

Analysis Sample

I introduced three restrictions to the data to construct my analysis sample. First, I restricted the base-year wave sample to respondents who participated in the 2004 follow-up wave, which reduces the sample from 15,244 to 15,236 students.[7] Second, I restricted the sample to students who were enrolled in public school, because private schools are not required to comply with the same state accountability requirements (N = 11,961). Third, I restricted the sample to respondents whose math teacher completed a questionnaire and have valid information for the primary explanatory variable, *math teacher expectations.* The resulting analytic sample contains 8,945 students enrolled in 561 schools.

[6] A state is defined as an exit exam state if, as of the 2001-02 academic school year, it required the graduating class of 2003-04 to pass an exit exam.
[7] Of the 35 base-year wave respondents removed from the sample, 21 were deceased and another14 were institutionalized at the time that the first follow-up was conducted.

For the models where the outcome variable is drawn from the first follow-up (i.e., models where the outcome variable is either twelfth grade math test or math test score gains), I restricted the sample to respondents who were enrolled in the twelfth grade and in the same school as in the base-year wave (N = 6,776 students enrolled in 555 schools). Finally, for the models where the outcome variable is one of the three enrollment status variables (i.e., graduate on time, held back a grade, or dropped out), I restricted the sample to students using enrollment status information drawn from the second follow-up (N = 8,611 students enrolled in 561 schools). Finally, I use best-subset regression imputation for all missing information and, because these are non-random samples of public school students, I use adjustment weights so that the findings may be interpreted as representative of public school students (see Appendix A for a discussion of the steps used for imputation and weight construction).

High School Versus Elementary School Students

The No Child Left Behind (NCLB) policy requires that states test all students annually in grades 3 through 8 and at least once again in high school. For this dissertation, I employ high school-level data rather than elementary-level data for several reasons. First, scholars have directed more attention to the association between accountability interventions and high school student performance, largely because there are more consequences for poor performance at the high school level, especially in terms of the transition to the labor market and post-secondary schooling. The emphasis on the transition beyond high school is grounded in the era of minimum competency testing, which I discuss in greater detail in Chapter 4. Second, I focus on high school students because research suggests that elementary school children are extremely impressionable to teacher expectations (Brenner and Mistry 2007; Rosenthal and Jacobson 1968), and therefore teacher effects may be inflated. High school students, on the other hand, are less impressionable because they have adjusted their expectations as they have moved through school.

Construct Validity of State Accountability Characteristics

How can I be sure that the accountability system characteristics I use are valid measures that capture the nature of each state's testing system? Research on accountability systems is accumulating rapidly, and with each new study scholars learn more about the problems associated with identifying causal effects. Prior studies typically use one of two approaches. The first is a pre-test/post-test study design to identify achievement gains associated with the implementation of accountability systems. The second examines the relationship between accountability characteristics (e.g., report cards or exit exams) and student achievement, where coefficients are interpreted as the direct effect of accountability testing on student achievement. While both approaches offer estimates of the achievement gains associated with accountability systems, they do not address the heterogeneity of accountability systems, and must therefore be interpreted with caution. Further complicating matters is the lack of available data to effectively analyze accountability interventions. Many studies match accountability characteristics to data drawn from national datasets, such as the National Assessment of Education Progress (NAEP), while others use the single case-study method to examine the achievement gains associated with implementation. In either case, the accountability literature focuses on how testing directly affects students, therein bypassing teachers.

This dissertation expands research on accountability systems by using a different source of data and by examining a different set of characteristics. When implementing accountability systems, states must decide what it is that they want to know. By way of modeling the type of tests administered (e.g., NRTs, CRTs, or

PATs), I am able to construct conclusions about how the *type* of information that states gather impacts teacher expectations. For example, NRTs are used to compare students to peers, while CRTs are used as indicators of how well students master core curriculum. PATs have similar goals, but are more complex estimates of student performance. I also examine the impact of testing rigor and exit exams in addition to types of tests on the student-teacher relationship.

CHAPTER 3

TEACHER EXPECTATIONS AND STUDENT ACHIEVEMENT: REVISITING A CLASSIC QUESTION

In this chapter, I address a classic question in the sociology of education: Do teachers influence achievement by communicating expectations to students? I first introduce the foundations of expectations research in both sociology and educational psychology. I then discuss the teacher expectations literature more narrowly and introduce a heuristic model of the determinants of teacher expectations and student achievement that integrates psychological and sociological models. Then, using the Education Longitudinal Study of 2002, I address two research questions. First, what factors predict teacher expectations of students' future educational attainment? Second, net of these predictors, how strong are teacher expectations of student educational attainment, where attainment is a measure of math test scores and enrollment status in 2004?

THE LEGACY OF *PYGMALION IN THE CLASSROOM*

Robert Rosenthal and Lenore Jacobson's (1968) *Pygmalion in the Classroom* was the first real-life psychology experiment that tested the self-fulfilling prophecy for student achievement. In their study, Rosenthal and Jacobson administered an intelligence test called the "Harvard Test of Inflated Acquisition" to all children enrolled in kindergarten through grade 5 in a low-income public school. The test was administered as a pre-test to measure student ability, but teachers were led to believe that it was a strong predictor of future "spurts" of learning (Rosenthal and Jacobson 1968:66). Rosenthal and Jacobson then selected a random sample of students and told the teachers these students were expected to be "late bloomers" who would make larger-than-normal achievement gains over the next year.

When students were retested at the end of the year, Rosenthal and Jacobson found that the late bloomers showed greater achievement gains than their peers. A more careful analysis revealed that achievement gains were largest among the first and second grade bloomers, which the authors attributed to the fact that younger children are more sensitive to teachers' influence. Rosenthal and Jacobson also explored behavioral differences and found that teachers were more approving of the bloomers, especially when they showed signs of achievement gains. In contrast, teachers responded negatively when the non-bloomer students' achievement gains were larger than anticipated. In other words, teachers responded positively when their expectations were confirmed (i.e., high achieving bloomers) and negatively when they were disconfirmed (i.e., high achieving non-bloomers). Rosenthal and Jacobson argued that teachers adjusted their behavior toward students according to their new expectations, thus producing a self-fulfilling prophecy. Rosenthal and Jacobson concluded:

> … we may say that by what she said, by how and when she said it, by her facial expressions, postures, and perhaps by her touch, the teacher may have communicated to the children of the experimental group that she expected improved intellectual performance. Such communications together with possible changes in teacher techniques may have helped the child learn by changing his self concept, his expectations of his own behavior, and his motivation, as well as his cognitive style and skills. (1968:180)

The *Pygmalion in the Classroom* findings were received with both curiosity and speculation. Immediately following its publication, both social psychologists and educational psychologists embarked upon a rigorous investigation to assess the validity of Rosenthal and Jacobson's findings. Educational and social psychologists approached the *Pygmalion* findings with different perspectives: educational psychologists were more skeptical of the power of teacher expectations, whereas social psychologists embraced the self-fulfilling prophecy (Jussim and Harber 2005).

In their extensive review of the previous 35 years of research, Jussim and Harber (2005) present a convincing case that the *Pygmalion* effects are genuine, arguing that previous research was unable to replicate Rosenthal and Jacobson's (1968) controversial findings because of the popularity of the study and because of a failure to accurately replicate *Pygmalion's* study design. They also argue that, firstly, the self-fulfilling prophecy *does* exist, but that the size of the effect of teacher expectations on student achievement tends to be small and is more likely to decrease rather than increase over time (Jussim 1989; Jussim and Eccles 1992; Smith, Jussim, and Eccles 1999). Second, they argue that self-fulfilling prophecies are more likely to occur among socially stigmatized groups because the stereotypes to which they are associated are, by definition, typically inaccurate. Teachers' perceptions of white and black students differed not because of the associated stereotypes, but rather because they were closely associated with past achievement (Jussim, Eccles, and Madon 1996). Nevertheless, Jussim and Harber argue that these differences in perceptions are indicative of a slight stereotype bias. Third, they acknowledge the association between teacher expectations and student achievement, but warn that making conclusions about whether teacher expectations actually increase (or decrease) student intelligence is beyond available evidence. And fourth, they argue that teacher expectations predict student achievement because they are accurate assessments of student achievement, which is the perspective maintained by the educational psychologists (Jussim 1989; Jussim and Eccles 1992; Jussim et al. 1996). In sum, teacher expectancy effects are important predictors of student achievement, but not necessarily self-fulfilling in that they cause large achievement gains, as theorized by Merton (1948).

THE WISCONSIN MODEL: THE INFLUENCE OF PARENTS, TEACHERS, AND PEERS COMPARED

Sociologists have been interested in the student-teacher relationship since at least the mid-1940s. One stream of research, best exemplified by Sewell, Haller, and Portes (1969), integrates social psychological and social structural explanations into a model of achievement that gives primacy to the importance of social influence. Figure 3.1 presents the educational attainment portion of the model of Sewell and colleagues (hereafter the “Wisconsin model”). As shown in the figure, a student’s prior academic performance is determined by mental ability and factors that generate the association between socioeconomic status and mental ability. The Wisconsin model posits that student aspirations are not just a function of student characteristics, but of the perceptions of parents, teachers, and peers, which is in turn affected by academic performance and socioeconomic status. Parents, teachers, and peers’ influence (what they call “significant others’ influence”) are expectations about future educational attainment that are communicated to students through social interactions, which then influence students’ own academic aspirations and, in turn, their future educational and occupational attainment.

Immediately following its publication, the Wisconsin model was criticized for being too simplistic and underwent many revisions (see Morgan 2005 for a review). The biggest issue was the generalizability of the theoretical model, given that the data used were restricted to a sample of farm boys drawn from the Wisconsin Longitudinal

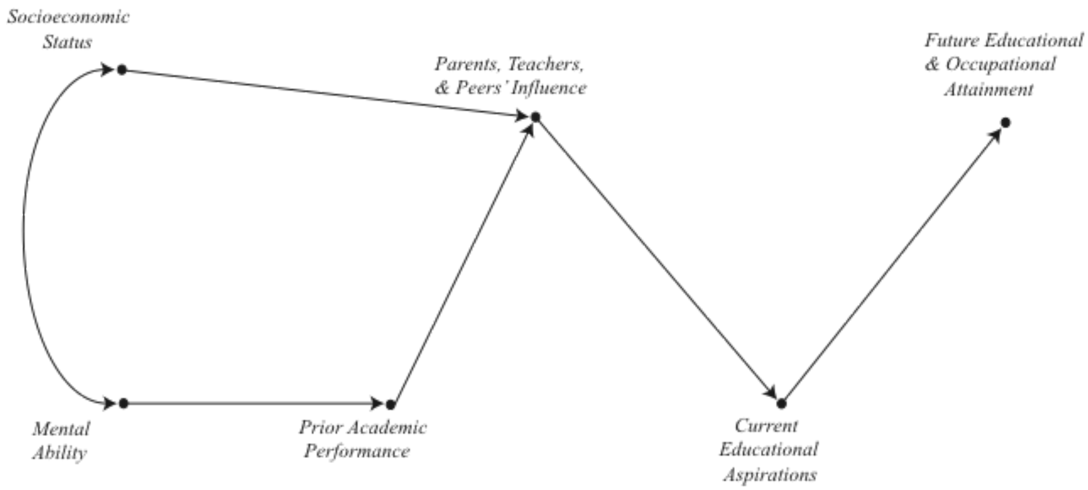


Figure 3.1: Nested Path-Model of the Educational Attainment Portion of the Wisconsin Model
Source: Sewell, Haller, and Portes (1969)

Study of 1957. The Wisconsin model hypothesis was later tested using both a broader sample of students and longitudinal data (Sewell, Haller, and Ohlendorf 1970), with much the same results. Although the authors amended the model by adding three additional paths, they still endorsed the ideas of the original Wisconsin model. After reviewing the Wisconsin model research, Sewell and Hauser (1980) concluded that a path could be drawn from each of the predictors of significant other's influence directly to educational and occupational achievement. Whether communicated verbally or via the subtleties of social interactions, parents, teachers, and peers are powerful agents who have the ability to influence individuals' own educational and career goals. In fact, perceptions may also be influenced by shared cultural beliefs, which lead to biased self-assessments of ability, therein channeling individuals into different career paths (Correll 2001).

Significant others' influence on individual students varies according to relationship to the student. Parents, for example, play the role of authority figures whose influence is oftentimes motivated by their own experiences and consequential aspirations for their children (Lareau 2003). Peers, on the other hand, act as role models to whom individuals compare themselves. Preceding the Wisconsin model by only a few years, James S. Coleman's (1961) *The Adolescent Society: The Social Life of the Teenager and its Impact on Education* is one of the first examples of how much weight peer influence has on students. He states,

> … our adolescents today are cut off, probably more than ever before, from the adult society. They are still oriented toward fulfilling their parents' desires, but they look very much to their peers for approval as well. Consequently, our society has within its midst a set of small teen-age societies, which focus teen-age interests and attitudes on things far removed from adult responsibilities, and which may develop standards that lead away from those goals established by the larger society. (Coleman 1961:9)

Although *The Adolescent Society* was published nearly sixty years ago, Coleman's theoretical framework continues to influence the ways in which scholars think about the importance of social influence, and how it is partially determined by the context in which individuals interact.

Recent research explores the ways in which peers influence each other academically. Crosnoe, Cavanagh, and Elder (2003), for example, find that students whose close friends value school or are good students are more likely to value school themselves. However, not all students respond similarly to their peers. In their study of the relationship between same-sex friendships and achievement, Riegle-Crumb, Farkas, and Muller (2006) find that young women are more responsive than young men to same-sex friends' academic achievement. That is, same-sex friends' achievement increases the likelihood of enrolling in advanced math and science courses for young women, but not for young men. Peer influence—and parent and teacher influence to a lesser extent—continues to be an important resource at even the higher education level (Bank, Slavings, and Biddle 1990). These studies suggest that students gravitate toward peers with similar interests, and that such ties also act as a source of motivation.

Teachers' influence, on the other hand, is restricted to the classroom setting, where teachers monitor students closely and direct them as needed. As a result, teacher expectations are influenced by their observations of student effort and performance, which they then communicate to students via constructive feedback and in social interactions with students. Teachers who feel responsible for student learning interact with students more frequently, thus increasing learning opportunities (Lee and Smith 1996).

Compared to parents and peers, teacher influence differs because teachers are responsible for motivating students to meet defined achievement expectations for which students are rewarded with grades and feedback. The more positively students perceive their teachers (e.g., teachers were caring and provided positive feedback) the more likely they are to be academically engaged (Hallinan 2008). Although teacher expectations are not positively associated with students liking school, students are perceptive of their teachers' behavior toward them (Hallinan 2008).

So far, I have introduced the psychological and sociological foundations of the expectations literature. Next, I discuss findings from the teacher expectations research more narrowly and present a heuristic model of teacher expectations that builds upon previous research in that it suggests that teacher expectations are also a function of context.

The Student-Teacher Relationship and the Importance of Context

Sociological research has examined the student-teacher relationship extensively, especially in the context of the black-white test score gap. Findings from this research show that teachers are more likely to perceive black students as being less academically engaged and more disruptive than their white peers (Ainsworth-Darnell and Downey 1998), which may be a function of white teacher bias (Downey and Pribesh 2004). Some researchers also claim that teachers are not only more likely to have less confidence in black students' ability to learn difficult material (Prime and Miranda 2006) but also more likely to attribute learning to external forces rather than student ability (Wiley and Eskilson 1978) and spend less time and effort instructing black and Hispanic students than white students (Porter et al. 1988). Teachers do not need to interact with students to construct expectations. In fact, they may associate student characteristics such as race or social class with social stereotypes. Something as simple as having a stereotypically black name can trigger teachers to lower their expectations (Figlio 2005). If teachers' differential treatment is not necessarily a function of student ability, how much does teacher influence matter? In order to answer this question, it is first important to understand what factors influence teacher expectations.

Teacher expectations for a given student are primarily driven by that student's background and ability (Alexander, Entwisle, and Bedinger 1994; Ferguson 2000; Harvey and Slatin 1975; Jussim et al. 1996). Teachers may also associate higher or lower ability with status characteristics, such as student race or gender (Correll and Ridgeway 2003). Teachers who adhere to such stereotypes are particularly harmful influences because their biased assessments in turn obstruct students' own aspirations (Ferguson 2000; Shepardson and Pizzini 1992). However, assuming that teachers are not averse to amending their assessments of students, the more often student performance exceeds teachers' expectations, the more likely teachers will revise their expectations such that they are consistent with a student's average performance. As a result, teacher expectations are more likely to be accurate assessments of student achievement (Brophy and Good 1974; Jussim 1989; Jussim and Eccles 1992; Jussim et al. 1996).

In the classroom, teachers evaluate students by comparing them to their peers and by considering their own experiences. For instance, teachers' own race may influence their perceptions of students, although evidence is mixed. Ehrenberg, Goldhaber, and Brewer (1985) argue that the match between student and teacher race and gender affects how teachers perceive students, but the student-teacher race match it is not associated with how much students actually learn. Babad (1985), on the other hand, argues that teachers' ethnicity does not bias expectations of students. More recently, Downey and Pribesh (2004) examined the student-teacher race match and found evidence that black teachers at the elementary and middle school levels tend to rate black students more favorably than white teachers rated white students. Downey and Pribesh argue that white teachers' more negative perceptions of black students is evidence of teacher bias rather than the black students exhibiting behaviors consistent with the oppositional culture argument because the findings are consistent for both elementary and middle school students.

Teacher quality may also affect the ways in which teachers and students interact. Teachers who have the proper training and certification are more sufficient at identifying and adapting to student needs (Hansen 1988), which in turn is associated with increased student learning (Bidwell and Kasarda 1975). Darling-Hammond, Berry, and Thoreson (2001) on the other hand, argue that certification does not matter because most teachers have qualifications that resemble certification. Moreover, teachers with advanced degrees are not always the most dedicated to their jobs (Riehl and Sipple 1996). In their analysis of the effects of licensure tests, Angrist and Guryan (2008) find no evidence of increased teacher quality. If anything, they find that state certification requirements are associated with an increase in salary. Given that the students in most need of such attention are more likely to be low-income and black or Hispanic and are also more likely to attend schools with less qualified teachers (Jordan, Mendro, and Weerasinghe 1997), state certification requirements may actually steer better teachers away from the students in most need.

The social dynamics of the student-teacher relationship are further complicated by the fact that teacher expectations are also a function of the environment in which teachers observe students. For example, the school size literature argues that larger schools are better equipped to address student needs because they can sort students into achievement groups (Lee and Loeb 2000; Lee and Smith 1995, 1997). At the same time, low performing students in small schools benefit from higher average expectations *because* their schools are small and therefore are less likely to sort students according to performance. Within the classroom, teachers interact differently with low- and high-expectancy students, which contributes to the achievement gap (Cooper and Tom 1984). Teachers tend to direct more attention at and provide more positive feedback to high-expectancy students (Brophy and Good 1970; Cooper 1979; Dusek 1975), while they give less attention and more critical feedback to low-expectancy students (Cooper 1979). Such differential treatment is particularly harmful when students adopt their teachers' expectations as their aspirations, especially if teachers do not update their expectations when they are disconfirmed (Jussim 1989).

Expectations are also manifested at the community level (Kozol 1991). Schools surrounded by poverty are more likely to have high rates of at-risk students, who are also more likely to be black and Hispanic. Teachers struggle to address behavioral issues, thus leaving less time to complete lesson plans. As a result, students fall behind their more affluent counterparts. Even if teachers aspire to make a difference in students' lives, their ambitions cannot compete with cultural expectations, and they lower their expectations accordingly.

The more that research uncovers about the complexity of the student-teacher relationship the more apparent it is that teachers' perceptions of students are a function of their academic environment (Prime and Miranda 2006). In order to understand how teachers influence students, one must first identify the student characteristics as well as the teacher and school characteristics that influence teacher expectations.

An Integrative Model of Teacher Expectations and Student Achievement

Figure 3.2 integrates social psychological and sociological models of attainment into a heuristic model of the observed determinants of teacher expectations and student achievement. In reality, the student-teacher relationship is much more complex, with many more causes and causal pathways that are not represented in the figure. This complexity will be reflected in the models presented in the results section; Figure 3.2 shows only the most crucial relationships.

In Figure 3.2, math teacher expectations and student achievement are a function of student characteristics (e.g., prior academic performance, race, and socioeconomic status), teacher characteristics (e.g., teachers' race, gender, and teaching experience), and school characteristics (e.g., percent of students who are minority, eligible for free lunch, and enrolled in the college preparatory track). [We can for now ignore the dashed line between math teacher expectations and past academic performance.] Teacher expectations, which are a function of student characteristics, school characteristics, and teacher characteristics, directly influence future academic performance. In addition to teacher expectations, future academic performance is a direct function of student characteristics and past academic performance. School characteristics indirectly influence student achievement by way of teacher expectations and past academic performance.

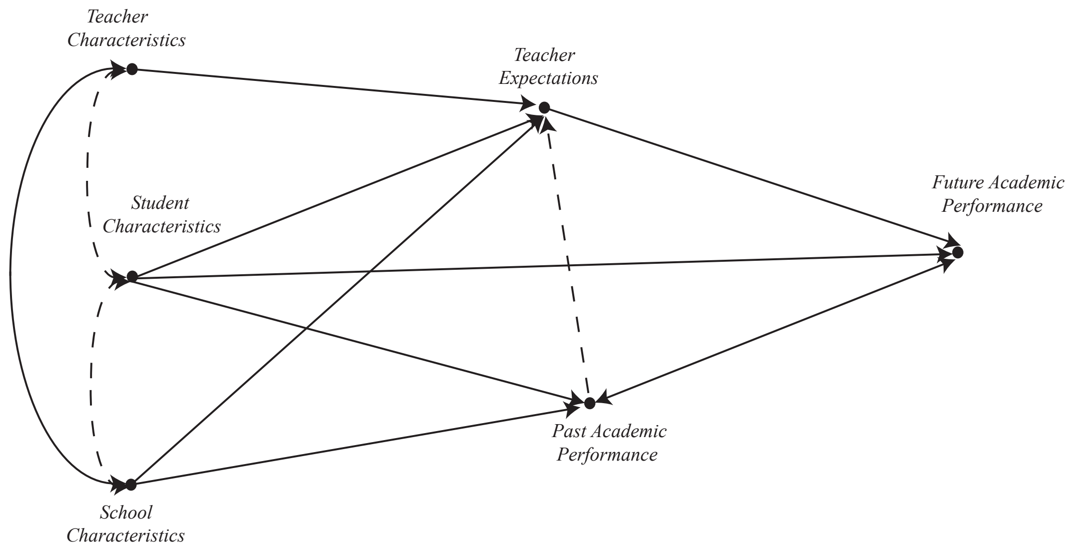


Figure 3.2: A Heuristic Model of the Observed Determinants of Teacher Expectations and Student Achievement

Even though Figure 3.2 suggests that teachers' influence on students is more complex than as presented in the Wisconsin model in Figure 3.1, the same ideas apply. Teachers use information about students to construct expectations, which they then communicate to students. In the model, teachers form initial impressions of students using knowledge about students, but also take into account their own experiences and the context in which they interact (i.e., school characteristics). Both teacher and school characteristics influence teacher expectations to the extent that teachers compare students to their peers. Teachers' initial expectations influence the ways in which they interact with students, which in turn influence student behavior (presented as past academic performance in Figure 3.2). Student behavior is also influenced by students' own characteristics as well as their school characteristics. Note that there is no direct causal path from teacher characteristics to student behavior. Instead, teacher characteristics influence student behavior through teachers' initial expectations. With respect to the psychological framework, the model also has a learning aspect to it, such that the more frequently teachers interact with students the more frequently they update their expectations (Braun 1976).

Before introducing the results, I would like to first define how teacher expectations are operationalized in this dissertation. The public discourse used to describe teacher expectations does not make a clear distinction between teachers' short-term (i.e., daily, weekly, or yearly) performance expectations and forecasts of future educational attainment. Whereas the former refer to performance on specific tasks, the latter are predictions of educational attainment.

For this dissertation, I define teacher expectations as a teacher's expectations for a given student's future educational attainment. I focus on teachers' forecasts of students' educational attainment for two reasons. First, the data employed in this dissertation do not offer information that allows me to model the short-term association between teacher expectations and student performance. They do, however, provide a measure of teachers' futuristic expectations. In particular, for the base-year wave of the Education Longitudinal Study (ELS), a teacher questionnaire was administered a math and English teacher for each student participant. Teachers were asked, "How far in school do you expect this student to get?" Responses range from "Less than high school graduation only" to "Will obtain a Doctorate, professional degree, or other advanced degree (Ph.D., M.D., etc.). Second, although the discourse used in education policy conflates the two types of expectations, it tends to focus more attention on teachers' expectations of future educational attainment, in part because of the emergence of the "college for all" culture that developed near the end of the twentieth century.

Next, I test the implications of the heuristic model using nationally representative data. In order to fully understand the association between math teacher expectations and student achievement, I first identify the predictors of teacher expectations. In these models, I identify how much student, teacher, and school characteristics influence teacher expectations. Net of the predictors, I then estimate the effect of math teacher expectations on student achievement and enrollment status. Before I present the results, I introduce the variables used for the analysis.

RESULTS

Table 3.1 presents the means and standard deviations for the primary variables of interest, all of which were drawn from the ELS. The primary explanatory variable, *math teacher expectations*, and the key predictor variables were all drawn from the 2002 base-year wave, at which point the students were enrolled in the tenth grade. The outcome variables—i.e., math test scores and 2004 enrollment status variables—were drawn from the 2004 and 2006 follow-up waves, but also include tenth grade math test scores which was drawn from the 2002 base-year wave. [See Chapter 2 and Appendix A for a more detailed description of the data.]

The student characteristics include race, gender, the five measures of socioeconomic status (i.e., mother's education, father's education, mother's occupational prestige score, father's occupational prestige score, and family income), family structure, academic track, and four variables that capture students' academic past (i.e., learning disability, ever enrolled in remedial math, ever held back, and ever had a behavior problem). In addition to these student characteristics, I use ninth grade GPA as an indicator of past academic performance. The teacher characteristics were drawn from the math teacher questionnaire and include teacher's race, gender, age, a dichotomous variable indicating whether teachers have their teaching certification, and the number of years of teaching experience. The school-level variables are all the common characteristics normally used in similar studies and include urbanicity, region, school size, the number of full-time teachers, and variables capturing the

Table 3.1: Means and Standard Deviations of Primary Variables of Interest

	Mean	S.D
Achievement Test Scores		
IRT estimated number right (10th grade)	41.96	14.14
IRT estimated number right (12th grade)	46.67	15.41
Gain Score (12th – 10th grade IRT estimated number right)	4.49	6.40
Enrollment Status in 2004		
Dropout	.05	
Retained	.02	
Graduated on time	.93	
Math Teacher Expectations (in years)	14.75	2.10
Female	.50	
Race (White is the reference category)		
Black	.15	
Hispanic	.16	
Asian	.04	
Native American	.01	
Multiracial	.04	
Family Background		
Mother's education (in years)	13.45	2.33
Father's education (in years)	13.59	2.59
SEI score of mother's occupation in 2002 (GSS 1989 coding)	44.98	12.88
SEI score of father's occupation in 2002 (GSS 1989 coding)	44.14	11.68
Family income (natural log)	10.60	1.13
Two-parent family	.75	
Academic Past (as reported by parent)		
Learning disability	.13	
Ever in remedial math	.10	
Ever held back	.14	
Ever have behavior problem	.09	
Academic Track (Academic is the reference category)		
General	.40	
Vocational	.12	
Math Teacher Characteristics		
Female	.55	
Black	.06	
Hispanic	.05	
Asian	.03	
Native American	.01	
Multiracial	.02	
Age (in years)	42.60	11.16
Certified degree or higher	.53	
Teaching experience (in years)	14.62	10.49
Urbanicity (Suburban is the reference category)		
Urban	.28	
Rural	.21	
Region (Midwest is the reference category)		
Northeast	.18	
South	.35	
West	.23	

	Mean	S.D
School Characteristics		
School enrollment	1477.51	843.76
Full-time teachers	83.32	42.89
Percent minority	35.85	30.65
Percent free-lunch	23.16	18.88
Percent college prep	55.24	29.39
Percent general track	51.27	34.40
Percent vocational track	18.16	20.48
Percent remedial math	6.94	9.24
9th Grade GPA	2.66	.82

Source: Education Longitudinal Study (2002, 2004, and 2006 waves)
Notes: The data are weighted by the NCES post-stratification weight (BYSTUWT) multiplied by the inverse probability of having non-missing information for the primary explanatory variable, math teacher expectations. The twelfth grade math test score and math gain score models are weighted by the base-year adjustment weight multiplied by the inverse probability of being enrolled in the twelfth grade and being in the same school at the time of the 2004 wave. The enrollment status variables are restricted to base-year wave respondents whose 2004 status was known. For these variables, the data are weighted by the base-year adjustment weight multiplied by the inverse probability of remaining in the sample. For all predictor variables and tenth grade math test score, *N* = 8,945 students enrolled in 561 schools. *N* = 6,776 students enrolled in 555 schools for the twelfth grade and math gain score models, and *N* = 8,611 students enrolled in 561 schools for the enrollment status variables.

percent minority students, free-lunch recipients, students enrolled in remedial math, college prep, general track, and vocational track students.

The Predictors of Teacher Expectations

In a simple world, teachers would base their expectations of students on academic performance only. We know from the extant research that this is not the case. Table 3.2 presents five models that predict math teacher expectations. Models 1 to 3 identify the student, teacher, and school characteristics determinants of math teacher expectations. Model 1 includes student characteristics only, whereas Model 2 adds teacher characteristics and Model 3 adds school characteristics. Models 3b and 4, which I discuss later, are included as supplementary models that test the robustness of the findings from Models 1 to 3. All non-dichotomous variables are centered around their means, so that the constants may be interpreted as the average years of schooling teachers expect students to complete.

The Model 1 coefficients suggest that teachers' assessments of students are sensitive to student characteristics. Take, for example, student race and gender, neither of which is an indicator of actual student ability. Girls are expected to complete an estimated .22 years of schooling more than boys, which is not surprising given that girls tend to be more compliant in the academic setting. In comparison, student race is a stronger predictor of teacher expectations. Teachers expect black and Hispanic students to complete an estimated .55 and .24 years of schooling less than white students, respectively. Asian students, by contrast, are expected to complete .74 years more than white students. Although the sizes of the coefficients for race and gender are quite small, they suggest that social stereotypes may influence teachers' assessments of students (Correll and Ridgeway 2003). The estimated coefficients also suggest that students from stable, affluent families are held to higher standards while students who are disadvantaged in some way are held to significantly lower standards. Together, the Model 1 coefficients imply that teachers' predictions of future educational attainment are sensitive to students' background.

Model 2 suggests that teacher characteristics are weak predictors of teacher expectations, net of the student characteristics. Black and Hispanic teachers have slightly higher average expectations, at .50 and .55, respectively, than white teachers. Moreover, adjusting for teacher race mediates student race, thus offering support for previous research indicating that teacher race matters (Downey and Pribesh 2004; Ehrenberg et al. 1985). Teachers who are certified have slightly higher expectations of students, suggesting that the more training teachers have the better equipped they are to adjust to students' needs and are therefore slightly more optimistic. Despite the fact that they are weak predictors of teacher expectations, teacher characteristics should not be ignored, as is evidenced by the slight shift in the sizes of the coefficients for student race.

Table 3.2: Math Teacher Expectations Regressed on Student, Teacher, and School Characteristics

	Model 1	Model 2	Model 3	Model 3b	Model 4
Female	.22	.20	.20	.19	-.02
	(.04)	(.04)	(.04)	(.04)	(.04)
Race (White is the reference category)					
Black	-.55	-.65	-.79	-.75	-.32
	(.09)	(.08)	(.09)	(.09)	(.08)
Hispanic	-.34	-.44	-.55	-.63	-.31
	(.09)	(.09)	(.09)	(.09)	(.08)
Asian	.74	.69	.58	.63	.42
	(.10)	(.10)	(.11)	(.11)	(.09)
Native American	-1.11	-1.04	-.95	-.83	-.52
	(.22)	(.23)	(.23)	(.27)	(.23)
Multiracial	-.19	-.22	-.28	-.28	-.09
	(.10)	(.10)	(.10)	(.11)	(.10)
Family Background					
Mother's education (in years)	.10	.10	.09	.08	.05
	(.01)	(.01)	(.01)	(.01)	(.01)
Father's education (in years)	.10	.10	.10	.09	.05
	(.01)	(.01)	(.01)	(.01)	(.01)
SEI score of mother's occupation in	.01	.01	.01	.01	.01
2002 (GSS 1989 coding)	(.00)	(.00)	(.00)	(.00)	(.00)
SEI score of father's occupation in	.00	.00	.00	.00	.00
2002 (GSS 1989 coding)	(.00)	(.00)	(.00)	(.00)	(.00)
Family income (natural log)	.10	.10	.09	.10	.08
	(.02)	(.03)	(.03)	(.03)	(.02)
Two-parent family	.12	.13	.14	.14	-.02
	(.06)	(.06)	(.06)	(.06)	(.05)
Past History (as reported by parent)					
Learning disability	-.90	-.90	-.91	-.90	-.61
	(.08)	(.07)	(.07)	(.07)	(.07)
Ever in remedial math	-.48	-.46	-.45	-39	-.30
	(.08)	(.08)	(.08)	(.07)	(.07)
Ever held back	-.81	-.81	-.83	-.82	-.52
	(.08)	(.08)	(.08)	(.07)	(.08)
Ever have behavior problem	-.78	-.77	-.77	-.76	-.33
	(.09)	(.09)	(.09)	(.09)	(.09)
Academic Track (Academic is the reference category)					
General	-.75	-.74	-.69	-.70	-.34
	(.05)	(.05)	(.05)	(.05)	(.05)
Vocational	-.74	-.74	-.73	-.73	-.40
	(.08)	(.08)	(.08)	(.08)	(.07)
Teacher Characteristics					
Female		.17	.14	.11	.08
		(.06)	(.6)	(.05)	(.05)
Black		.50	.41	.47	.54
		(.11)	(.11)	(.13)	(.11)
Hispanic		.55	.46	.38	.36
		(.20)	(.17)	(.16)	(.17)
Asian		.23	.23	.43	.09
		(.16)	(.17)	(.19)	(.17)

	Model 1	Model 2	Model 3	Model 3b	Model 4
Native American		-.45	-.36	-.79	-.37
		(.33)	(.27)	(.29)	(.33)
Multiracial		.25	.19	.40	.18
		(.19)	(.19)	(.19)	(.17)
Age (in years)		-.01	-.01	-.01	-.01
		(.00)	(.00)	(.00)	(.00)
Certified degree or higher		.13	.11	.07	.04
		(.05)	(.06)	(.06)	(.05)
Teaching experience (in years)		.01	.01	.01	.00
		(.01)	(.00)	(.00)	(.00)
Urbanicity (Suburban is the reference category)					
Urban			.12		.15
			(.08)		(.07)
Rural			-.09		-.05
			(.06)		(.07)
Region (Midwest is the reference category)					
Northeast			.26		.42
			(.08)		(.08)
South			-.02		.01
			(.07)		(.07)
West			-.21		-.32
			(.09)		(.09)
School Characteristics					
School size			-.00		-.00
			(.00)		(.00)
Full-time teachers			.00		.00
			(.00)		(.00)
Percent minority			.00		.01
			(.00)		(.00)
Percent free lunch			-.00		-.00
			(.00)		(.00)
Percent college prep			.00		.00
			(.00)		(.00)
Percent general			-.00		-.00
			(.00)		(.00)
Percent vocational			-.00		-.00
			(.00)		(.00)
Percent remedial math			-.00		.00
			(.00)		(.00)
9th Grade GPA					1.19
					(.03)
Constant	15.40	15.20	15.35	15.29	15.09
R^2	.28	.29	.30	.38	.45

Source: Education Longitudinal Study (2002 wave)
Notes: The data are weighted by the NCES post-stratification weight (BYSTUWT) multiplied by the inverse probability of having non-missing information for the primary explanatory variable, math teacher expectations. *N* = 8,945 students enrolled in 561 schools.

Teachers' sensitivity to student race becomes even more apparent when the social context variables are added to the model (see Model 3). The results presented in Model 3 also suggest that school characteristics do influence teacher expectations, insofar as the context in which teachers and students interact provides teachers with information about student ability and performance. However, the small coefficients make it difficult to discern how much context really matters. To test the robustness of the findings of Model 3, Model 3b adjusts for school fixed-effects. School fixed-effects models include a dummy for all schools but one, thereby adjusting for all between-school variability. The remaining coefficients may therefore be interpreted as the average within-school relationships between the predictor variables and teacher expectations. Adjusting for school fixed-effects has very little impact on the student and coefficients for teacher characteristics, thus suggesting that the context in which teachers interact with students does in fact matter.

The models discussed so far offer evidence that student characteristics are the primary source of information on which teachers base their expectations. In the academic setting, it is not uncommon for teachers to discuss students with other teachers. Teachers use this information to construct initial expectations, which then influence how they perceive students. Even though teachers update their expectations of students over time, knowledge of prior academic performance prepares them for what to expect, and in some sense sets a baseline expectation prior to any direct information the teacher observes about a student.

Research in the Wisconsin model tradition typically adjusts for prior academic performance, a strategy I follow in Model 4. The ELS data do not offer test scores prior to the tenth grade, but they do offer retrospective reports of student grades. Model 4 adds ninth grade GPA as a measure of prior academic performance to the Model 3 adjustment variables. Compared to Model 3, including past performance decreases the coefficients for student characteristics: the coefficient for gender is explained away entirely, and the coefficients for race, the four measures of academic past, and academic track are all reduced substantially. The positive residual associations support the argument that teacher expectations are influenced by more than just student achievement

So far, the results show that student, teacher and school characteristics are important predictors of teacher expectations. The extent to which each set of predictor variables explains teacher expectations, however, is unknown because variance-explained is a complex indicator of relative importance (as indicated in the many critiques of the Coleman Report; Coleman et al. 1966). The amount of variance that each set of predictor variables explains is a function of assumptions about causal order and model specification (e.g., student, teacher, and school characteristics separately or in unison) as well as measurement error.

One way to hone in on causal order is to estimate the upper and lower bound estimates of variance explained by each set of characteristics. The upper bound estimate is the proportion of the variance explained by the most simple model in which coefficients for only one set of characteristics are estimated. For example, the upper bound estimate of the variance explained by the student characteristics is estimated from the following model:

$$TE_i = \beta_0 + \beta_1 STU_i + e, \tag{3.1}$$

where TE_i represents teacher expectations for individual i, β_0 is the constant, and STU_i is a vector for student characteristics. The lower bound estimate, or the unique contribution to the variance explained, is the difference between variance explained by the fully specified model (i.e., Model 3):

$$TE_i = \beta_0 + \beta_1 STU_i + \beta_2 TCH_i + \beta_3 SCH_j + e, \qquad (3.2)$$

and the variance explained without the variables of interest:

$$TE_i = \beta_0 + \beta_1 TCH + \beta_3 SCH_j + e, \qquad (3.3)$$

where TCH_i is a vector for teacher characteristics for student *i* and SCH_j is a vector for school characteristics in school *j*. The upper and lower bound estimates give the range of the amount of the variance explained by each set of predictor variables.

The top panel of Table 3.3 presents bounds for the amount of variance explained by the student, teacher, and school characteristics. The lower and upper bounds indicate that student characteristics have the greatest influence on teacher expectations, explaining 24 to 28 percent of the variance. School characteristics explain an additional one to six percent of the variance, and teacher characteristics explain about one percent. Collectively, these estimates offer support for past research, which claims that teacher expectations are primarily a function of student characteristics, but that other factors also influence how teachers perceive students.

Now consider the bottom panel, which presents the amount of variance explained by each set of predictor variables when prior academic performance is included as a predictor of teacher expectations (i.e., Model 4 in Table 3.2). Recall that including prior academic performance reduced the associations between the student characteristics and teacher expectations. Adjusting Equations (3.2) and (3.3) to include prior academic performance pushes the lower bound estimate down for student characteristics from 24 to six percent of the variance. Meanwhile, prior academic performance explains between 15 and 34 percent of the variance. It is still unclear

Tale 3.3: Upper and Lower Bound Estimates of the Percent Variance-Explained by Student, Teacher, and School Characteristics as well as Past Academic Performance

A. Model 3 Specification

	Lower Bound	Upper Bound
Student Characteristics	.24	.28
Teacher Characteristics	.01	.01
School Characteristics	.01	.06

B. Model 4 Specification

	Lower Bound	Upper Bound
Student Characteristics	.06	.28
Teacher Characteristics	.01	.01
School Characteristics	.02	.06
9^{th} Grade GPA	.15	.34

Source: Education Longitudinal Study (2002 wave)
Notes: See Table 3.2 notes.

how to interpret this because, chronologically, academic performance follows the student predictors. Still, the lower and upper bound estimates indicate that teacher expectations are primarily a function of student characteristics and prior academic performance.

In sum, student characteristics and past academic performance are the strongest predictors of teacher expectations, which is to be expected. The fact that student characteristics explain about one-quarter of the variance suggests that teachers strongly consider a student's social background in estimating his or her future educational attainment. They also offer evidence that teacher characteristics mediate teacher expectations, albeit modestly, which implies that teachers' assessments of students are a function of teachers' own characteristics and years of teaching experience. Furthermore, Models 3 and 4 in Table 3.2 show that school characteristics also affect math teacher expectations because they shape the social environment in which teachers observe and interact with students. Next, I examine how much teacher expectations matter for student achievement, net of these predictors.

Teacher Expectations and Student Achievement

In the remainder of this chapter, I explore the relationship between math teacher expectations and student test scores and enrollment status in 2004, net of the predictor variables identified in the previous section. Table 3.4 presents the coefficients for math teacher expectations from regression models predicting tenth grade math test scores, twelfth grade math test scores, and math test score gains (i.e., 12^{th} – 10^{th} grade). [The full models are available in Tables B1, B2, and B3 in Appendix B.] Each panel presents coefficients from six regression models predicting math test scores. The first column presents the bivariate association between teacher expectations and Models 1 to 4 include the same adjustment variables as Models 1 to 4 in Table 3.2. It should be noted that math teachers did not know how students performed on the tenth grade math test at the time they completed the teacher questionnaire. Therefore, the findings can be interpreted as the association of teachers' expectations on student performance.

The coefficients presented in the top panel suggest that math teacher expectations are positively associated with tenth grade math test scores. The Unadjusted Model indicates that each additional year of school that teachers expect students to complete is associated with an increase of 3.98 points on the tenth grade math test. The association decreases substantially to 2.77 when adjusting for student characteristics (see Model 1), and remains at this level in models that further adjust for teacher and school characteristics, respectively (see Models 2 and 3). The Model 3b coefficient offers evidence that, net of student, teacher, and school characteristics, each additional year of schooling that teachers expect students to complete is associated with a 2.87-point gain in tenth grade math test scores. Further adjusting for prior academic performance reduces the association by 25 percent, from 2.83 to 2.12 (see Model 4). The adjusted effect of teacher expectations is quite large. For

Table 3.4: Math Teacher Expectations Coefficients from Regression Models Predicting 10th and 12th Grade Math Test Scores and Math Test Gains

	Outcome Variable: 10th Grade Math Test					
	Unadjusted Model	Model 1	Model 2	Model 3	Model 3b	Model 4
Math Teacher Expectations	3.98	2.77	2.82	2.83	2.87	2.12
	(.08)	(.08)	(.08)	(.08)	(.07)	(.08)
	Outcome Variable: 12th Grade Math Test					
	Unadjusted Model	Model 1	Model 2	Model 3	Model 3b	Model 4
Math Teacher Expectations	4.69	3.28	3.35	3.34	3.42	2.52
	(.12)	(.12)	(.11)	(.11)	(.10)	(.12)
	Outcome Variable: Math Gain (12th – 10th Grade Math Test)					
	Unadjusted Model	Model 1	Model 2	Model 3	Model 3b	Model 4
Math Teacher Expectations	.58	.49	.49	.48	.49	.32
	(.06)	(.06)	(.06)	(.06)	(.06)	(.07)
Student Characteristics		✓	✓	✓	✓	✓
Teacher Characteristics			✓	✓	✓	✓
School Characteristics				✓		✓
School Fixed-Effects					✓	
9th Grade GPA						✓

Source: Education Longitudinal Study (2002 and 2004 waves)
Notes: The tenth grade math test score models are weighted by the NCES post-stratification weight (BYSTUWT) multiplied by the inverse probability of having non-missing information for the primary explanatory variable, math teacher expectations. The twelfth grade math test score and math gain score models are weighted by the base-year adjustment weight multiplied by the inverse probability of being enrolled in the twelfth grade and being in the same school as the base-year wave at the time of the 2004 wave. *N* = 8,945 students enrolled in 561 schools for the tenth grade math test score models, and *N* = 6,776 students enrolled in 555 schools for the twelfth grade and math gain score models.

example, consider a student whose math teacher expects him or her to graduate from high school only versus completing a Bachelor's degree. The student expected to graduate from college will score an estimated 8.48 points (i.e., 2.12 * 4 = 8.48) more on the tenth grade math test than a student who is expected finish with a high school diploma only, or .60 standard deviations (i.e., 8.48 / 14.14 = 60).

The results discussed so far suggest that math teacher expectations are strongly associated with tenth grade math test scores. How lasting or predictive of future attainment are teacher expectations? To answer these questions, I offer two additional analyses in which I model the association between math teacher expectations in the tenth grade and both twelfth grade math test scores and math test score gains.

Consider the middle panel, which presents the coefficients for math teacher expectations from models predicting twelfth grade math test scores. The sizes of the coefficients are slightly larger but follow a similar pattern as the coefficients predicting tenth grade math test scores. For example, the bivariate association is slightly larger, at 4.69 versus 3.98, thus suggesting that math teacher expectations in the tenth grade are associated with achievement gains between the tenth and twelfth grades. Adjusting for student, teacher, and school characteristics reduces the size of the coefficient from 4.69 to 3.28 in Model 1, and to 3.35 and 3.34 in Models 2 and 3, respectively. Even after adjusting for prior academic performance, the association remains quite strong at 2.52. The bottom panel, which presents the results associated with test score gains, further gives evidence that math teacher expectations are associated with achievement gains over time.

Together, these findings offer convincing support for the claim that teacher expectations matter for student outcomes. According to the Model 3 results presented in Table 3.4, net of student, teacher, and school characteristics, students whose math teachers expect them to graduate from college have tenth grade math test scores that are 11.32 points higher than their peers who are expected to finish with a high school diploma (i.e., 2.83 * 4 = 11.32), or .80 standard deviations (i.e., 11.32 / 14.14 = .80). By the twelfth grade, the same difference increase by another two points (i.e., .48 * 4 = 1.92), or .30 standard deviations (i.e., 1.92 / 6.40 = .30). Although small, the estimated differences between the tenth and twelfth grades suggest that students benefit from having teachers who hold them to higher standards. Next I examine the relationship between teacher expectations and the likelihood of students graduating on time, being retained (i.e., held back a grade), and dropping out of school between 2002 and 2004.

Teacher Expectations and Student Enrollment

Table 3.5 presents the predicted changes in the probability of graduating on time, being held back a grade, and dropping out of school by the spring of 2004, separately for students expected to graduate from high school (and go no further in school) versus those expected to graduate with a Bachelor's degree or higher.[8] To calculate the changes in the probabilities, I first estimated the likelihood of each outcome. Then, using the estimated models, I estimated the discrete changes in the probability of students expected to graduate (1) from high school by 2004 and (2) with a Bachelor's degree or higher, holding all covariates at their means. The estimated probabilities presented may thus be interpreted as the differences in the average probability of students graduating on time, being held back, or dropping out of school by spring 2004. Each panel presents the discrete changes in the probability for five models: the Unadjusted Model includes only math teacher expectations; Model 1 includes student characteristics; Model 2 includes teacher characteristics; Model 3 includes school characteristics; and Model 4 adds prior academic performance.

Consider first the differences in the probability of students graduating on time. The Unadjusted Model indicates that students whose teachers expect them to graduate from high school and not go to college have a .831 probability of graduating by 2004. In comparison, students expected to graduate from college or higher have a .979 probability of graduating on time, for a difference of .148. The difference between the two groups decreases substantially from .148 to .082 when adjusting for student

[8] For each panel, students who graduated early were treated as being on track to graduate on time. Removing them from the analysis does not significantly affect the odds for either outcome.

Table 3.5: Predicted Changes in the Probability of Graduating on Time, Being Held Back, and Dropping Out of School by Spring 2004 for Students Expected to Graduate from High School Only Versus Students Expected to Graduate with at Least a Bachelor's Degree

	Outcome Variable: Graduate On Time		
	Graduate from High School Only	Graduate from College or Higher	Pr[College – High School]
Unadjusted Model: Math teacher expectations	.831	.979	+ .148
Model 1: Unadjusted Model + student characteristics	.902	.984	+ .082
Model 2: Model 1 + teacher characteristics	.976	.995	+ .019
Model 3: Model 2 + school characteristics	.973	.995	+ .022
Model 4: Model 3 + 9th grade GPA	.987	.997	+ .010
	Outcome Variable: Held Back (Out of Grade 12)		
	Graduate from High School Only	Graduate from College or Higher	Pr[College – High School]
Unadjusted Model: Math teacher expectations	.041	.009	- .032
Model 1: Unadjusted Model + student characteristics	.026	.008	- .018
Model 2: Model 1 + teacher characteristics	.010	.003	- .007
Model 3: Model 2 + school characteristics	.011	.004	- .007
Model 4: Model 3 + 9th grade GPA	.007	.003	- .004
	Outcome Variable: Dropout		
	Graduate from High School Only	Graduate from College or Higher	Pr[College – High School]
Unadjusted Model: Math teacher expectations	.127	.013	- .114
Model 1: Unadjusted Model + student characteristics	.067	.009	-. 058
Model 2: Model 1 + teacher characteristics	.015	.002	- .013
Model 3: Model 2 + school characteristics	.017	.003	- .014
Model 4: Model 3 + 9th grade GPA	.007	.001	- .006

Source: Education Longitudinal Study (2002 and 2006 waves)
Notes: The coefficients and standard errors for math teacher expectations are available in Table B4 of Appendix B. The enrollment status variables are restricted to base-year wave respondents whose 2004 status was known, and therefore the sample is reduced to 8,611 students enrolled in 561 schools. For these variables, the data are weighted by the base-year adjustment weight multiplied by the inverse probability of remaining in the sample.

characteristics (see Model 1), to .022 when adjusting for teacher and school characteristics (see Model 3), and to .010 when adding prior academic performance (see Model 4).

The middle panel presents the predicted changes in the probability of students being held back a grade (i.e., out of grade 12 in the spring of 2004). The differences in the predicted probabilities suggest that teacher expectations have less influence on the likelihood of students being retained a grade. For example, the difference between the Unadjusted Model estimates indicates that students expected to graduate from high school only are .032 times more likely to be held back a grade than their expected college-graduate peers. However, this association is explained away almost entirely when the student, teacher, and school predictor variables and prior academic performance are included in the model (see Models 1, 2, 3, and 4), which shows that teacher expectations have less influence on student retention rates.

The bottom panel of Table 3.5, by contrast, suggests that teacher expectations have slightly more impact on the likelihood of students dropping out of school. The predicted probability of dropping out of school for students expected to graduate from high school only and go no further is .127, compared to .013 for students expected to at least complete college. Again, the difference in the discrete changes in the probability of the two outcomes decreases substantially from .114 (see Unadjusted Model) to .058 when adjusting for student characteristics (see Model 1), to .014 when adjusting for teacher and school characteristics (see Model 3), and to .006 when adjusting for past academic performance (see Model 4).

Together, the three panels presented in Table 3.5 offer additional evidence in support of the importance of teachers' influence on students. Although taking into account the predictor variables substantially reduces the difference in the predicted probabilities, the fact that the difference is not reduced entirely suggests that teacher expectations do predict on-time graduation and dropout, although not retention. For struggling students, merely having a teacher who has high expectations may mean the difference between graduating and falling behind.

DISCUSSION

Scholarship in sociology and educational psychology identifies teachers as integral actors in the learning process, but the magnitude of their influence has remained a source of debate (Jussim and Harber 2005; Ferguson 1998, 2003). In this chapter, I examined the relationship between math teacher expectations and student achievement. I first identified the predictors of teacher expectations, and then modeled the relationship between math teacher expectations and student achievement. The results for the first analysis suggest that a wide range of characteristics predict teacher expectations, but that teacher expectations are primarily a function of student characteristics. The results from the second analysis show that math teacher expectations strongly predict student achievement, whether achievement is a measure of math test scores or the likelihood of students staying on track and graduating on time. This finding holds true even after the common causes of math teacher expectations are taken into account. Moreover, in these two analyses, I do not attempt to make any claims about the causal order of the predictor variables. Rather, I estimate the extent to which the predictors identified in decades of research on teacher expectations explain math teacher expectations. Although the magnitude of the effect varies across outcomes, the results suggest that teacher expectations are strong predictors of student achievement.

Current education policy calls upon teachers to raise student achievement by holding all students accountable to meet a uniform standard. The argument maintains that teachers who hold socially stigmatized groups to lower standards are responsible for the size of achievement gaps. Therefore, pressuring teachers to raise their expectations will raise student achievement for these groups, and in turn narrow the achievement gaps. In the next chapter, I examine how teachers respond to state-level pressure to raise their expectations, and whether accountability interventions do, in fact, raise student achievement.

CHAPTER 4

TEACHER EXPECTATIONS AND STUDENT ACHIEVEMENT IN THE ERA OF SCHOOL ACCOUNTABILITY

This chapter examines how teachers respond to state-level pressure to raise their expectations for students. In the remainder of this chapter, I briefly discuss the history of the testing movement, followed by a review of findings from the literature on accountability systems. I then present a heuristic model of student achievement that incorporates accountability interventions into the model presented in Chapter 3. Using the Education Longitudinal Study data matched to the state-level accountability dataset that I constructed, I address three research questions: (1) How do teachers respond to accountability interventions? (2) Do teachers communicate these expectations to students? and (3) What are the consequences for student learning for low- and high-performing students? Finally, I discuss the results in the context of current education policy.

REVIEW OF THE HISTORY OF TESTING IN THE UNITED STATES

Tests have been used since as early as the 1840s (Resnick 1982), but it was not until the middle of the twentieth century that they were used for purposes other than routine monitoring of student performance. Since then, the testing movement has undergone several transitions, each of which was driven by a change in the primary motivation for testing. In this section, I briefly describe the major waves of testing at the primary and secondary school levels, paying particular attention to how tests have shifted from being used as mere assessments of student performance to assessments with consequences (for extensive reviews, see Grodsky, Warren, and Felts 2008; Hamilton 2003; Koretz 2008a; Linn 2000, 2008).

The first major transition was sparked by James B. Conant (1953)'s argument that tests should be used to identify students eligible for higher education. He maintained that using tests for admissions decisions would make the college application process more competitive by opening doors to students who would normally not consider pursuing a post-secondary education. Colleges and universities responded accordingly, relying on norm-referenced tests (NRTs) for admission decisions. NRTs assess student knowledge of basic skills and score students relative to the average for the reference populations. This "norm-referencing" allows colleges and universities to gauge the quality of the students they accept relative to the national pool of college applicants. Today, NRTs are still commonly used tests. For example, NRTs are used to evaluate student performance throughout elementary and secondary school (e.g., ITBS/ITED, SAT-9, and CAT-7), and as entrance exams for college (e.g., SAT and ACT) and higher education programs (e.g., GRE, MCAT, and LSAT). Despite their widespread use, NRTs are generally considered a secondary alternative to criterion-referenced tests, because of the fit of the test to the content (see below).

The second wave of testing was initiated by the enactment of the Elementary and Secondary Education Act of 1965 (ESEA). Nearly a decade after the seminal *Brown v. Board of Education* ruling that declared "separate but equal" unconstitutional, the federal government responded to the lack of progress made by states in closing the achievement gaps between black and white students. ESEA defined the federal government as the primary overseer of the highly decentralized public education system. ESEA included many programs, each of which was targeted toward improving the expectations and quality of education for targeted groups. One such program, Title 1, required that states administer minimum competency tests (MCTs) to low-income students (i.e., students who were more likely to be non-white and live in the South) to ensure that they could demonstrate knowledge of basic skills.

To educators, MCTs did not pose a problem as long as student performance was above the minimum threshold. Policymakers, on the other hand, were concerned that they were too basic and therefore "let educators off the hook."

Whether or not MCTs were vigorous assessments of student performance, the severity of educational inequality became clear after 35 states used minimum competency tests by the end of the 1970s (Koretz 2008a). For the first time educators and policymakers had evidence that allowed them to assess the quality of the public education system. Moreover, state departments of education acknowledged that the lack of structure contributed to achievement gaps and used the opportunity to define clear content standards (i.e., "what" students should know) and performance standards (i.e., "how much" students should know). Several states also attached sanctions to test scores, thus marking the beginning of the accountability movement.

The third wave of testing was initiated by the publication of *A Nation At Risk* (National Commission on Excellence in Education 1983), a federally mandated review of the public education system. The report assessed student performance at both the national and international levels and concluded that the American school system was failing. Policymakers were appalled, and called for increased accountability. The report was the first to make international comparisons, which only heightened the call made by policymakers and American observers for educators to monitor student achievement more closely.

Throughout the 1980s, most states made a concerted effort to hold schools and students accountable for meeting defined achievement expectations. The heightened sense of responsibility resulted in an increased reliance on test usage, thus legitimizing tests as a means of measuring student performance. It also called for more rigorous tests that would provide the information necessary to assess overall student performance relative to defined content standards. States responded by replacing

MCTs with criterion-referenced tests (CRTs). Unlike MCTs, CRTs are aligned to state curricula and use proficiency-level cut-scores (e.g., "advanced," "proficient," and "basic") rather than a single baseline for performance expectations. During this wave, states expanded the population of students they tested to include all students, rather than only low-income students authorized to be tested under Title 1 of the ESEA. As a result, educators and policymakers gained better knowledge of student performance relative to state standards. However, because CRTs are state-specific, student performance in one state cannot be compared to performance in another state.

By the late 1980s, several states took an additional step toward greater accountability by administering performance assessment tests (PATs), which are thought to be more challenging assessments of student knowledge than CRTs or NRTs. PATs measure similar information as CRTs, but rather than using multiple choice or true/false questions, they require students to present their work (e.g., proofs or written essays), therein allowing teachers to follow how students answer questions and identify where their logic differs. Despite the fact that PATs offer more detailed information about student performance, they are less common than NRTs and CRTs because they are difficult to grade and much more expensive to grade (Koretz 2008a).

Today, all states rely on some form of testing, using a mixture of NRTs, CRTs, or PATs. Although CRTs are, by far, the most commonly used test, many states administer additional tests as well. Prior to NCLB, there was no defined national framework on which to structure accountability systems; therefore, states that administered more than one type of test benefitted from having different types of evidence of student performance. Furthermore, because no one test is superior to others for all purposes, the more information that states gathered about students, the better equipped they were to make adjustments to their education systems. All states tested students at both the elementary and secondary school levels, but elementary students were tested more frequently. In addition to determining which tests to use, states also determined which subjects they tested students. English and math were the most frequently tested subjects, however, many states also tested student performance in science and/or social studies.

NCLB introduced a school-based accountability system that increased achievement standards and the frequency of testing, but also attached sanctions to poor student performance. Under NCLB, all states are required to adopt more rigorous content and achievement standards, and tests are used to determine whether students meet these performance expectations. All states are required to test students annually in math and reading in grades 3 through 8 and at least once in high school. Since NCLB was enacted, both science and social studies tests have been added to the testing regime. Test scores are used to set annual achievement targets for the "annual yearly progress" report, which is used to help all students reach the NCLB target whereby all students are "proficient" by 2014. Thus, NCLB built upon the lessons learned from the previous 30 years of testing. In the next section, I review the research on accountability systems. I also discuss the complications associated with examining the impact of testing on teachers' responses to state-level pressure to raise their expectations of students, and subsequently on student achievement.

Assessing the Impact of Testing on Student Achievement

The fundamental goal of accountability interventions is to motivate teachers to teach and students to learn. Tests purportedly assess whether this goal is met. Despite the fact that accountability tests have been used since the 1970s, it is unclear if they produce the intended outcomes. The lack of consensus of their effects is in part due to the lack of available data that allows scholars to study the heterogeneity of accountability systems, and in part to the variety of techniques used to model accountability systems (Linn 2008), some of which generate different estimates.

To date, research primarily focuses on the direct impact of accountability systems on student achievement. Some research suggests that accountability expectations may increase test scores, but that their effect tends to differ by students' race and socioeconomic status (Carnoy and Loeb 2002; Hanushek and Raymond 2005; Roderick, Jacob, and Bryk 2002). Some researchers claim that merely introducing accountability interventions increases student test scores (Grissmer and Flanagan 1998; Ladd 1999; Texas Agency 2000), grades (Roderick et al. 2002), and work effort (Roderick and Engel 2001). Even if test scores increase when accountability systems are first introduced, the association between accountability expectations and test scores tends to weaken over time (Linn, Graue, and Sanders 1990).

More recently, Fuller, Wright, Gesicki, and Kang (2007) show that overall test score growth has on average decreased since NCLB was passed, they also show, however, that states have different rates of test score growth, which they argue is a function of different achievement standards before and after NCLB was enacted. Specifically, many states already had high performance standards (e.g., 70th percentile) before NCLB was enacted (Linn 2003). The only difference is that, before NCLB, states were not required to attach sanctions to performance standards. As a result, states that had initially high achievement standards prior to NCLB made fewer achievement gains because their achievement standards were already high. By contrast, states that had low achievement standards prior to NCLB showed greater achievement gains *because* they were forced to raise their achievement expectations and also because there were real-life consequences attached to poor performance (Fuller et al. 2007). It is also possible that achievement gains are a function of states simultaneously increasing graduation requirements (Schiller and Muller 2003). In other words, accountability interventions may be associated with test scores, but they appear to be most effective when they are first introduced or accompanied by a revision of achievement expectations.

Within the literature on accountability systems, scholars have also examined how sanctions attached to test scores affect student outcomes. Some claim that sanctions such as exit exams are associated with higher achievement gains (Winfield 1990) while others find no evidence of an association (Grodsky, Warren, and Kalogrides 2009). However, Jacob (2001) cautions that any association between sanctions and achievement can be explained by adjusting for student characteristics. Adjusting for school context, Carnoy and Loeb (2002) find that student achievement gains are greatest in states with stronger accountability systems, but the type of accountability (i.e., report card or sanctions) is not associated with significantly different achievement gains (Hanushek and Raymond 2004).

Further complicating matters, it is unclear whether achievement gains are a result of low-performing students having been sorted into special education or retained the year prior to an important assessment (Amrein-Beardsley and Berliner 2003; Figlio and Getzler 2006; Jacob 2005; see also Hanushek and Raymond 2003). Some researchers find evidence that low-performing schools protect themselves from falling below state standards by sorting low-performing students into alternative tracks where they are exempt from accountability expectations (Darling-Hammond 1991; Haney 2000). Schools that use such a strategy may do more harm to students than intended, because the learning standards are lower for students enrolled in special education. Similarly, students who are retained the year prior to an important assessment do not necessarily benefit from the additional year of preparation (Roderick, Nagaoka, and Allensworth 2005), and are also more likely drop out of school (Hauser 2004).

Research examining the association between accountability interventions and dropout or retention rates is also unclear. Some evidence suggests that accountability testing decreases the likelihood of high school graduation, especially among traditionally disadvantaged groups (Carnoy and Loeb 2002; Darling-Hammond 2003; Warren, Jenkins, and Kulick 2006). Evidence on whether sanctions (e.g., high school exit exam requirements) increase or decrease dropout rates is also unclear (Jacob 2001; Ladd 1999; Lillard and DiCicca 2001).

After 40 years of research, it thus is still unclear whether accountability interventions produce the intended outcomes. One of the biggest shortcomings of the research on accountability systems is the lack of rich data to examine such questions. Much of the extant research focuses on single case studies, such as Chicago (e.g., Jacob 2001; Roderick and Engel 2002; Roderick et al. 2002; Roderick et al. 2005), Texas (e.g., Deere and Strayer 2003; Haney 2000; Ladd 1999; Texas Education Agency 2000), or Kentucky (e.g., Koretz and Barron 1998). These analyses offer rich information about each case, but insofar as they are predictive, the results cannot be extrapolated easily to other cities or states. Some studies match accountability characteristic data to nationally representative data (e.g., the National Assessment of Educational Progress, also known as "NAEP") and estimate the achievement gains associated with accountability interventions (e.g., Carnoy and Loeb 2002; Grodsky et al. 2009; Hanushek and Raymond 2004). However, as Olson (2005) shows, the NAEP achievement standards tend to be higher than state standards, and as a result, studies that use these data tend to report lower rates of proficiency. Evidence also demonstrates that accountability test scores do not necessarily generalize to scores on other tests, such college entrance exams (Klein, Hamilton, McCaffrey, and Stetcher 2000; Koretz and Barron 1998), therein bringing into question the validity of what accountability tests measure. Scholars are still learning how to address the methodological complications association with such complex research questions and how to model state accountability characteristics (see Koretz 2008b for a discussion). In addition to these concerns, Raudenbush (2004) argues that studies often fail to discount competing hypotheses because they do not take into account prior academic performance and school context differences.

So far I have discussed findings from research that examines how students respond to accountability expectations, suggesting that accountability interventions may or may not raise student achievement. The research discussed thus far focus on the direct association between accountability interventions and student outcomes. Teachers, as mediators in the learning process, are overlooked, primarily because the data used do not include information about teachers. What research does exist tends to focus on teachers' responses to accountability interventions only. As a result, it is unclear whether the association between accountability interventions and student performance is a result of increased student awareness of accountability expectations, teachers responding to state-level pressure and more effectively communicating achievement expectations to students, or both. Next, I discuss what is known about how teachers respond to accountability interventions.

Teachers' Responses to Accountability Interventions

Teachers often oppose the use of accountability tests as a means of measuring student achievement because state-level standards interfere with their teaching agendas. Although tests are intended for use as direct measures of student achievement, they are also possible indicators of teacher effectiveness, and therefore may be threatening to teachers. The more pressure teachers feel to meet expectations, the more likely they are to adjust their teaching plans in order to prepare students for the tests, modifying their teaching strategies and content matter to spend more time on material they know will be on the tests (Abrams et al. 2003; Linn 1994; Taylor et al. 2003; see McDonnell 2004 for a review). For example, in their analysis of Colorado teachers, Taylor and colleagues (2003) found that teachers were quite responsive to changes in state-defined achievement expectations, especially when the subject matter was to be tested. Teachers aligned lesson plans according to the new content standards, which improved the overall level of instruction, but also admitted to setting aside instruction time in order to prepare students for the test. This affected the quality of students' education insofar as the more time teachers spent preparing students for subjects tested (e.g., math and English, the less time they spent on subjects not tested (e.g., science and social studies). Teachers may also respond by directing more attention to the students who are close to the cut point (Hamilton et al. 2007), or by adjusting student test scores in order to meet achievement cut-off points (Dewan 2010; Jacob and Levitt 2002). Overall, most teachers oppose the use of tests because tests introduce additional challenges to everyday teaching demands.

Even if teachers do not believe that tests are necessary to monitor student attainment, they are useful tools that, when used properly, help teachers identify students in need of additional help (Shields et al. 2004). Furthermore, attaching sanctions to tests introduces an incentive for teachers and students to perform well. Students' mastery of state-defined expectations is a matter of many factors, most of which rely on teachers making the necessary adjustments to instruction plans and communicating expectations to students such that they adopt them as their own.

A MODEL OF TEACHER EXPECTATIONS AND STUDENT ACHIEVEMENT IN THE ERA OF SCHOOL ACCOUNTABILITY

Although the fate of many of the control provisions of NCLB will be debated, accountability testing is likely to continue to be a prominent feature of education in the

United States. Despite the fact that tests have been used for several decades, much is to be learned about how to measure expectations and how to model them accordingly, especially given that testing can be stressful for both students and teachers (Jones et al. 1999; Steele 1997). Furthermore, while the extant research offers some insight as to how students and teachers respond to state accountability standards, very little is known about how teachers mediate the effects of these accountability expectations.

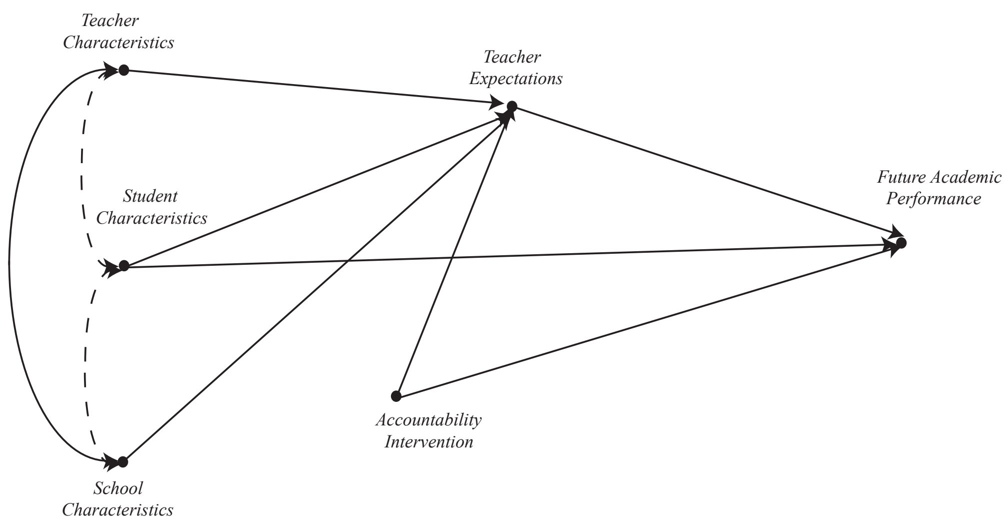


Figure 4.1: A Heuristic Model of the Observed Determinants of Teacher Expectations and Student Achievement, Adjusting for Accountability Interventions

Recall Figure 3.2 from Chapter 3, which presented a heuristic model of the observed determinants of teacher expectations and student achievement. In the model, teacher expectations affect student achievement directly as well as through a set of common causes (i.e., student, teacher, and school characteristics). Now consider how accountability interventions affect the association between teacher expectations and student achievement. The model presented in Figure 4.1 introduces two additional paths to the model presented in Figure 3.2: (1) from accountability interventions to teacher expectations and (2) from accountability interventions to future academic performance.[9] In the model both teacher expectations and accountability interventions affect student achievement directly. The model also allows teacher expectations to affect student achievement indirectly by way of accountability interventions.

In the next section, I introduce the state-level accountability data and discuss the ways in which state accountability interventions vary across states. Then, using the model presented in Figure 4.1, I address the following questions: (1) How do teachers respond to accountability interventions? (2) Do teachers communicate these expectations to students? and (3) What are the consequences for student learning for low- and high-performing students?

RESULTS

Characteristics of State Accountability Systems

Table 4.1 presents the 2001-02 accountability system characteristics for grades 9 to 12 for each of the 50 states and the District of Columbia (see Chapter 2 for a detailed description of the data and variable construction). The table highlights the complexity of state accountability systems. The first column identifies states that were early participants in the movement to use tests as a means of monitoring student performance. Compared to the early 1970s, when only a few states used standardized tests to monitor student performance, 36 states used tests by 1991. By the 2001-02 academic school year, all 50 states and the District of Columbia administered standardized tests to students, either to monitor student progress or as accountability measures. The next three columns identify the type of tests that states used, followed by whether states had an exit exam requirement. The final four columns present the number of times that states administered math, English, science, and social studies

[9] Note that although the model is a simplified version of Figure 3.2, the same ideas apply.

Table 4.1: Characteristics of 2001-02 State Accountability Systems by State

		Type of Test				Testing Frequency in High School				
	Early Adopter	PAT	NRT	CRT	Exit Exam	Math	English	Science	Social Studies	Test Index
Alabama				1	1	1	1			-.77
Alaska	1		1	1	1	2	2			-.21
Arizona			1	1		2	2			-.21
Arkansas			1	1	1	2	2		1	.04
California	1		1	1	1	3	3	3	3	1.81
Colorado	1		1	1		3	3	1		.60
Connecticut	1	1		1		1	1	1		-.52
Delaware	1		1	1		1	1	1	1	-.28
DC			1			3	3			.36
Florida		1	1	1	1	2	2	1		.04
Georgia	1			1	1	1	1	1	1	-.28
Hawaii	1		1	1		1	1			-.77
Idaho	1		1	1		3	3	2	2	1.33
Illinois	1		1	1		1	1	1	1	-.28
Indiana	1		1	1	1	1	1			-.77
Iowa	1		1			1	1	1		-.52
Kansas				1		1	1			-.77
Kentucky			1	1		2	4	1	1	.86
Louisiana	1		1	1	1	2	2	2	2	.76
Maine	1	1		1		1	1	1	1	-.28
Maryland	1			1	1	2	2	1	1	.28
Massachusetts	1			1	1	1	1			-.77
Michigan				1		1	1	1	1	-.28
Minnesota	1	1		1	1		1			-1.04
Mississippi	1			1	1	3	3	3	3	1.81
Missouri	1	1	1	1		1	1	2	1	-.04
Montana			1			1	1	1	1	-.28
Nebraska		1	1	1		1	1			-77
Nevada	1			1	1	2	2			-.21
New Hampshire				1		1	1	1	1	-.28
New Jersey	1			1	1	1	1			-.77
New Mexico	1			1	1	2	2	2	2	.76
New York	1			1	1	2	1		2	-.01
North Carolina	1	1	1	1	1	4	4	4	4	2.86
North Dakota			1			1	1			-.77
Ohio	1			1	1	1	1	1	1	-.28
Oklahoma	1			1			1		1	-.80
Oregon	1	1		1		1	1	1		-.52
Pennsylvania	1	1		1		1	2			-.48
Rhode Island	1			1		1	2	1		-.23
South Carolina	1			1	1	1	1			-.77
South Dakota			1			1	2	1	1	.01
Tennessee	1	1		1	1	1	2	1		-.23
Texas	1			1	1	2	2	1	1	.28
Utah	1		1	1		3	3	3		1.09

	Early Adopter	Type of Test			Exit Exam	Testing Frequency in High School				
		PAT	NRT	CRT		Math	English	Science	Social Studies	Test Index
Vermont	1	1		1		1	1	1		-.52
Virginia	1		1	1	1	4	2	3	4	2.03
Washington			1	1		2	2			-.21
West Virginia	1		1			3	3	3		1.09
Wisconsin	1		1	1		1	1	1	1	-.2
Wyoming			1	1		1	1			-.77

Source: State-Level Accountability Characteristics
Notes: See Chapter 2 for a discussion of the sources of information and variable construction.

tests to students between grades 9 and 12, and the last column presents an index of the frequency of testing. One component of the index is the subjects that were tested (i.e., math, English, science, and social studies). All states except Minnesota and Oklahoma tested students in math and English during the 2001-02 academic school year. Few states, by contrast, tested science and social studies. A second component of the index is testing frequency, which differs by subjects as well as from state to state.

Table 4.2 presents the state-level and student-level means and standard deviations for variables presented in Table 4.1 as they are matched to the ELS data. Looking at the state-level means, nearly three out of four states (or 72 percent) were early participants in the testing movement. CRTs are by far the most common type of test used (90 percent), and PATs are the least common type of test administered (22 percent), which is not surprising given that they are so costly to grade. Fifty-two percent of states use NRTs. Only 44 percent of states required students to pass an exit exam to graduate. On average, states administered between one and two tests in both English and math tests, and only about one test of science and social studies.

Table 4.2: State-Level and Student-Level Means and Standard Deviations for the Characteristics of 2001-02 State Accountability Systems

	State-Level		Student-Level	
	Mean	S.D	Mean	S.D.
Early Implementer	.72		.79	
PAT	.22		.23	
NRT	.52		.48	
CRT	.90		.97	
Exit Exam	.44		.64	
Frequency of Testing Index	-.00	.84	.17	.97
Frequency of Math Testing	1.60	.93	1.73	.97
Frequency of English Testing	1.70	.86	1.75	.88
Frequency of Science Testing	.96	1.03	1.12	1.14
Frequency of Social Studies Testing	.76	1.04	1.09	1.21

Sources: State-Level Accountability Characteristics and the Education Longitudinal Study (2002 wave)
Notes: The state-level means and standard deviations (N = 50) are not weighted. The student-level means (N = 8,945 students enrolled in 561 schools) and standard deviations are weighted by the NCES post-stratification weight (BYSTUWT) multiplied by the inverse probability of having non-missing information for the primary explanatory variable, math teacher expectations.

Do Accountability Interventions Influence Teacher Expectations?

Recall Table 3.2 in Chapter 3, which identified the student, teacher, and school level determinants of teacher expectations. The findings demonstrate that although teacher expectations are primarily a function of student characteristics, the context of schooling also matters. The current chapter extends this logic to argue that the social context of schools is, in part, a function of state-defined standards. Accordingly, the following analysis examines how accountability interventions influence teacher expectations, adjusting for the student, teacher, and school determinants of teacher expectations identified in Table 3.2.

Table 4.3 shows the coefficients for accountability interventions from models that regress math teacher expectations on accountability interventions and the adjustment variables. For the following analysis, each accountability characteristic is modeled separately. The first column of each panel presents the bivariate association between each accountability intervention and math teacher expectations. Model 1 adjusts for student and teacher characteristics while Model 2 adjusts for school characteristics only. Model 3 then adjusts for student, teacher, and school characteristics, and Model 4 further adjusts for ninth grade GPA. The models do not adjust for region (as this is superfluous when state characteristics are modeled). Consider the top panel, which examines whether math teacher expectations in states with a history of testing differ from those in states that only recently joined the accountability movement. The bivariate association suggests that math teacher expectations are slightly higher in states that were early adopters of accountability systems. The association, which is already quite weak, decreases from .16 to .11 when adjusting for student and teacher characteristics (see Model 1), and is explained away entirely when adjusting for school characteristics only (see Model 2). When ninth Grade GPA is included in the model (see Model 4), the size of the coefficient and standard error suggest that early implementation has no influence on teacher expectations.

The next three panels examine the association between the *type* of test administered and math teacher expectations. Recall that each test offers a unique assessment of student knowledge. PATs offer rigorous assessments of student ability, but are costly. NRTs assess more basic knowledge and provide scores that are scaled to the reference population. CRTs, the most common type of test, assess student knowledge of state-defined curricula, and therefore offer insight into student performance relative to defined achievement expectations.

The three panels in Table 4.3 suggest that there is no systematic pattern of association between the type of test used and math teacher expectations. For example, the bivariate association suggests that teachers in states that administer PATs have, on average, slightly higher teacher expectations. This weak association is explained away by adjusting for student and teacher characteristics (see Model 1) and school characteristics (see Model 2). By contrast, the bivariate association for NRTs suggests that they are associated with slightly lower math teacher expectations. The association

Table 4.3: Accountability Intervention Coefficients from Regression Models Predicting Math Teacher Expectations

	Unadjusted Model	Model 1	Model 2	Model 3	Model 4
Early Implementer	.16	.11	-.00	.01	.15
	(.09)	(.06)	(.08)	(.06)	(.06)
	Unadjusted Model	Model 1	Model 2	Model 3	Model 4
PAT	.17	.01	.09	.04	.05
	(.10)	(.06)	(.08)	(.06)	(06)
	Unadjusted Model	Model 1	Model 2	Model 3	Model 4
NRT	-.23	-.21	-.15	-.14	-.18
	(.08)	(.06)	(.07)	(.06)	(.06)
	Unadjusted Model	Model 1	Model 2	Model 3	Model 4
CRT	-.03	-.04	-.26	-.17	-.01
	(.20)	(.13)	(.18)	(.12)	(.13)
	Unadjusted Model	Model 1	Model 2	Model 3	Model 4
Exit Exam	-.03	-.03	.04	-.03	.05
	(.08)	(.06)	(.07)	(.06)	(06)
	Unadjusted Model	Model 1	Model 2	Model 3	Model 4
Testing Index	-.10	-.08	-.05	-.07	-.05
	(.05)	(.03)	(.04)	(.03)	(.03)
Student Characteristics		✓		✓	✓
Teacher Characteristics		✓		✓	✓
School Characteristics[a]			✓	✓	✓
9th Grade GPA					✓

Sources: State-Level Accountability Characteristics and the Education Longitudinal Study (2002 wave)
Notes: The data are weighted by the NCES post-stratification weight (BYSTUWT) multiplied by the inverse probability of having non-missing information for the primary explanatory variable, math teacher expectations. N = 8,945 students enrolled in 561 schools.
[a] Region is not included as an adjustment variable.

decreases only slightly from .23 to .21 when adjusting for student and teacher characteristics (see Model 1) and to .15 when adjusting for school characteristics (see Model 2). Not adjusting for anything, CRTs have no association with math teacher expectations. Once school characteristics are taken into account, the coefficient increases from -.03 (see Unadjusted Model) to -.26 (see Model 2), and to -.17 when student and teacher characteristics are added (see Model 3). However, the negative association is explained away entirely when 9th grade GPA is added to the model (see Model 4). Together, these three panels suggest that teachers do not respond to the type of information gathered.

In addition to using tests to monitor student performance, some states require that students pass an exit exam in order to graduate. The fifth panel, which presents the coefficients for exit exams, suggests that exit exams also have no impact on teacher expectations. The last panel examines whether testing frequency influences math teacher expectations. In theory, accountability tests offer teachers additional information upon which to evaluate students, and hence one might anticipate an association with teacher expectations. The coefficients for testing frequency demonstrate that this measure has no net impact on math teacher expectations.

The findings presented in Table 4.3 do not provide evidence that supports policy claims that teachers raise their expectations to accountability standards. If anything, they suggest that teachers lower their expectations of students.

Accountability Interventions, Teacher Expectations, and Student Achievement

Do accountability characteristics affect the student-teacher relationship? To address this question, I present two pieces of analysis. I first present findings from a model that allows both math teacher expectations and accountability interventions to predict student achievement directly. I then present findings from a model that includes the interaction effect between math teacher expectations and accountability interventions. Findings from this model reveal whether accountability interventions alter the association between math teacher expectations and tenth grade math test scores.

Table 4.4 shows results from models regressing tenth grade math test scores on each type of accountability intervention, math teacher expectations, and the adjustment variables. The models correspond to those presented earlier in Table 4.3. Following the Unadjusted Model, Model 1 adjusts for student and teacher characteristics while Model 2 adjusts for school characteristics only. Model 3 then includes the student, teacher, and school predictor variables, and Model 4 further adjusts for 9^{th} grade GPA. Similar to the analysis presented in Table 4.3, each characteristic of accountability testing is modeled separately. Finally, the top panel, which does not include any of the accountability intervention variables, is presented for reference purposes only.

Before discussing the association between accountability interventions and tenth grade math test scores, first consider how adding the accountability intervention variables to the models affects the direct effect of math teacher expectations on math test scores. Compared to the results presented in the top panel, the sizes of the coefficients for math teacher expectations do not change when the accountability intervention variables are included in the models, thus demonstrating that math teacher expectations and the measures of accountability interventions are largely unrelated.

Now consider the association between accountability interventions and tenth grade math test scores. The second panel presents the results for the models that estimate the association between early adoption and tenth grade math test scores. The coefficients suggest that math test scores are slightly higher in states that were early adopters. For example, the Unadjusted Model indicates that math test scores are .38 points higher in states that were early adopters. This association remains at about this level when adjusting for the student, teacher, and school predictor variables (see

Table 4.4: Coefficients for Math Teacher Expectations and Accountability Interventions from Regression Models Predicting 10th Grade Math Test Scores

	Unadjusted Model	Model 1	Model 2	Model 3	Model 4
Math Teacher Expectations	3.98	2.82	3.71	2.82	2.14
	(.08)	(.08)	(.08)	(.08)	(.09)

	Accountability Intervention: Early Implementer				
	Unadjusted Model	Model 1	Model 2	Model 3	Model 4
Math Teacher Expectations	3.97	2.82	3.71	2.82	2.13
	(.08)	(.08)	(.08)	(.08)	(.09)
Early Implementer	.38	.42	.38	.24	.71
	(.61)	(.42)	(.46)	(.42)	(.42)

	Accountability Intervention: Performance Assessment Test				
	Unadjusted Model	Model 1	Model 2	Model 3	Model 4
Math Teacher Expectations	3.97	2.82	3.71	2.82	2.14
	(.08)	(.08)	(.08)	(.08)	(.09)
PAT	1.24	.26	.03	.11	.20
	(.56)	(.44)	(.44)	(.42)	(.40)

	Accountability Intervention: Norm-Referenced Test				
	Unadjusted Model	Model 1	Model 2	Model 3	Model 4
Math Teacher Expectations	3.98	2.82	3.71	2.82	2.13
	(.08)	(.08)	(.08)	(.08)	(.08)
NRT	-.05	-.32	-.01	-.32	-.56
	(.52)	(.35)	(.41)	(.37)	(.36)

	Accountability Intervention: Criterion-Referenced Test				
	Unadjusted Model	Model 1	Model 2	Model 3	Model 4
Math Teacher Expectations	3.98	2.82	3.71	2.82	2.14
	(.08)	(.08)	(.08)	(08)	(.09)
CRT	-.43	1.05	1.04	1.23	1.66
	(1.68)	(1.27)	(1.37)	(1.25)	(.1.12)

	Accountability Intervention: Exit Exam				
	Unadjusted Model	Model 1	Model 2	Model 3	Model 4
Math Teacher Expectations	3.98	2.82	3.71	2.82	2.13
	(.08)	(.08)	(.08)	(.08)	(08)
Exit Exam	-.72	.86	.68	.94	1.19
	(.57)	(.39)	(.41)	(.37)	(.37)

	Accountability Intervention: Testing Index				
	Unadjusted Model	Model 1	Model 2	Model 3	Model 4
Math Teacher Expectations	3.97	2.83	3.72	2.83	2.14
	(.08)	(.08)	(.08)	(.08)	(.09)
Testing Index	-.26	.23	.40	.29	.31
	(.26)	(.20)	(.21)	(.19)	(.19)
Student Characteristics		✓		✓	✓
Teacher Characteristics		✓		✓	✓
School Characteristics[a]			✓	✓	✓
9th Grade GPA					✓

Sources: State-Level Accountability Characteristics and the Education Longitudinal Study (2002 wave)
Notes: See Table 4.3 notes.
[a] Region is not included as an adjustment variable.

Models 1, 2, and 3), but increases slightly to .71 when further adjusting for prior academic performance (see Model 4). In other words, math test scores tend to be slightly higher in states that were early to implement an accountability system. The next three panels present the results for PATs, NRTs, and CRTs, respectively. The coefficients show that the associations differ by test type. For example, the coefficient for the bivariate association for PATs is 1.24 (see Unadjusted Model). This association is explained away almost entirely when the student, teacher, and school predictor variables and prior academic performance are included in the model (see Models 1, 2, 3, and 4). NRTs, on the other hand, are associated with lower test scores, but only when student and teacher characteristics (see Models 1 and 3). By contrast, the bivariate association for CRTs is -.43, suggesting that math test scores are on average slightly lower in states that use CRTs (see Unadjusted Model). This association then becomes positive once the predictor variables are added to the model (see Models 1, 2, 3, and 4), which can be explained by the fact that states that use CRTs tend to have lower average socioeconomic status. The coefficients for both exit exams and the testing index follow a similar pattern as CRTs, where the bivariate associations are slightly negative, but become positive once student, teacher, and school characteristics are added to the model. Similar to CRTs, both exit exams and testing frequency are also associated with slightly higher math test scores once the adjustment variables are included in the model.

The results presented in Table 4.4 show that, except for NRTs, accountability interventions are associated with slightly higher math test scores. They also show that models that fail to adjust for student characteristics may in fact mask the association between accountability interventions and math test scores (Raudenbush 2004). The models presented in Table 4.4 assume that the association between math teacher expectations and math test scores does not differ according to compliance with accountability standards. However, it is possible that teachers adjust their expectations to accountability tests, and, in turn, teachers may more accurately predict students' future educational attainment. Table 4.5 shows results from models that allow math teacher expectations to differ according to accountability testing requirements. For this analysis, I regress tenth grade math test scores on each accountability intervention, math teacher expectations, the interaction between math teacher expectations and accountability interventions, and the adjustment variables.

Consider the second panel, which presents the estimates of whether teachers in states that were early to implement accountability standards differ from later adopters. Compared to the results presented in Table 4.4, including the interaction effect in the model does not affect the main effects for either math teacher expectations or early implementation. The interaction effect indicates that math teacher expectations in states that were early to adopt an accountability system are not more predictive of math test scores.

The next two panels, which present the results for PATs and NRTs respectively, also show that adjusting for the interaction effect does not affect the size

Table 4.5: Coefficients for Math Teacher Expectations, Accountability Interventions, and the Interaction Between Math Teacher Expectations and Accountability Interventions from Regression Models Predicting 10th Grade Math Test Scores

	Unadjusted Model	Model 1	Model 2	Model 3	Model 4
Math Teacher Expectations (MTE)	3.98	2.82	3.71	2.82	2.14
	(.08)	(.08)	(.08)	(.08)	(.09)

	Accountability Intervention: Early Implementer				
	Unadjusted Model	Model 1	Model 2	Model 3	Model 4
Math Teacher Expectations	4.00	2.83	3.72	2.84	2.14
	(.17)	(.14)	(.16)	(.14)	(.14)
Early Implementer	.37	.42	.38	.24	.71
	(.61)	(.43)	(.46)	(.42)	(.42)
*Early Implementer * MTE*	-.04	-.01	-01	-.03	-.01
	(.19)	(.16)	(.18)	(.16)	(.15)

	Accountability Intervention: Performance Assessment Test				
	Unadjusted Model	Model 1	Model 2	Model 3	Model 4
Math Teacher Expectations	3.89	2.80	3.64	2.79	2.12
	(.09)	(.09)	(.09)	(.08)	(.09)
PAT	1.21	.25	.00	.09	.19
	(.55)	(.44)	(.44)	(.42)	(.39)
*PAT * MTE*	.34	.13	.33	.15	.06
	(.18)	(.16)	(.18)	(.16)	(.15)

	Accountability Intervention: Norm-Referenced Test				
	Unadjusted Model	Model 1	Model 2	Model 3	Model 4
Math Teacher Expectations	3.95	2.75	3.67	2.74	2.09
	(.10)	(.09)	(.10)	(.09)	(.10)
NRT	-.04	-.31	-.01	-.32	-.56
	(.52)	(.35)	(.41)	(.37)	(.36)
*NRT * MTE*	.06	.14	.08	.16	.08
	(.16)	(.13)	(.15)	(.13)	(.13)

	Accountability Intervention: Criterion-Referenced Test				
	Unadjusted Model	Model 1	Model 2	Model 3	Model 4
Math Teacher Expectations	2.96	2.27	2.92	2.33	1.64
	(.42)	(.36)	(.36)	(.35)	(.31)
CRT	-.46	1.03	1.00	1.20	1.63
	(1.64)	(1.25)	(1.33)	(1.24)	(1.11)
*CRT * MTE*	1.04	.57	.81	.50	.51
	(.42)	(.36)	(.37)	(.35)	(.32)

	Accountability Intervention: Exit Exam				
	Unadjusted Model	Model 1	Model 2	Model 3	Model 4
Math Teacher Expectations	3.96	2.80	3.67	2.80	2.05
	(.13)	(.12)	(.12)	(.12)	(.12)
Exit Exam	-.72	.86	.68	.94	1.19
	(.57)	(.39)	(.41)	(.37)	(.37)
*Exit Exam * MTE*	.03	.04	.06	.04	.12
	(.17)	(.14)	(.15)	(.14)	(.13)
	Accountability Intervention: Testing Index				
	Unadjusted Model	Model 1	Model 2	Model 3	Model 4
Math Teacher Expectations	3.97	2.83	3.72	2.83	2.14
	(.08)	(.08)	(.08)	(.08)	(.09)
Testing Index	-.25	.23	.41	.30	.31
	(.26)	(.20)	(.21)	(.20)	(.19)
*Testing Index * MTE*	.08	.08	.10	.08	.08
	(.08)	(.07)	(.08)	(.07)	(.07)
Student Characteristics		✓		✓	✓
Teacher Characteristics		✓		✓	✓
School Characteristics[a]			✓	✓	✓
9th Grade GPA					✓

Sources: State-Level Accountability Characteristics and the Education Longitudinal Study (2002 wave)
Notes: See Table 4.3 notes.
[a] Region is not included as an adjustment variable.

of the coefficients for the direct effects. Math teacher expectations continue to be strong predictors of tenth grade math test scores, whereas PATs are associated with slightly higher test scores and NRTs are associated with lower test scores. The coefficients for the interaction effects indicate that math teacher expectations are slightly more predictive of math test scores in states that use PATs or NRTs.

Shifting focus to CRTs, the results presented in the fourth panel show that the relationship between math teacher expectations and student achievement is stronger in states that use CRTs. Recall from Table 4.4 that the base estimates for math teacher expectations and CRTs were 3.98 and -.43, respectively (see Unadjusted Model). The analogous results presented in Table 4.5 show that including the interaction effect in the model affects the size of the coefficient for math teacher expectations but not CRTs. For example, not adjusting for any of the predictor variables, the direct effect for math teacher expectations is 2.96 and the interaction effect is 1.04 (see Unadjusted Model), suggesting that math teacher expectations most strongly predict math test scores in states that use CRTs. The size of the interaction effect is reduced by half from 1.04 to .57 when adjusting for student and teacher characteristics (see Model 1), and only slightly to .81 when adjusting for school characteristics only (see Model 2). The results presented in Model 4 show that math teacher expectations more strongly predict math test scores in states that use CRTs, even after prior academic performance is taken into account.

Now consider the coefficients for exit exams. Many states require that students pass an exit exam in order to graduate, therein introducing motivation to do well. The direct effect for exit exams shows that math test scores are slightly higher in states that require students to pass an exit exam, suggesting that students are aware of the consequences of poor performance. The interaction effects, on the other hand, indicate that teacher expectations are not substantially higher in states that require students to pass an exit exam.

Whereas exit exams have consequences for poor performance, frequent testing is another means of closely monitoring students. In other words, teacher expectations in states that rely heavily on testing (e.g., North Carolina and California) may be more predictive of student achievement because they have more information about student performance. The findings presented in the bottom panel show that testing frequency is positively associated with tenth grade math test scores, as indicated by the slightly positive direct effects. However, similar to the results for exit exams, testing frequency has no impact on teachers' influence on student achievement.

The results presented in Tables 4.4 and 4.5 show that accountability interventions may in fact increase student achievement, but not necessarily because teachers are raising their expectations. First, accountability testing is associated with slightly higher math test scores, although the achievement gains tend to be quite small. Second, math teacher expectations continue to be strongly associated with math test scores. Third, the findings presented in Table 4.5 show that the strength of the association between math teacher expectations and math test scores differs according to accountability intervention status, thus suggesting that teachers use information from accountability tests to adjust their expectations for students. The pattern of the strength in the relationship between accountability interventions and math teacher expectations is strongest for CRTs, then PATs, NRTs, and exit exams.

Why do math teachers appear to adjust their expectations to CRTs but not any of the other characteristics of accountability interventions? Recall that the CRT movement began in response to increased demand for improved student performance, especially among low-income, minority children. Accordingly, states that use CRTs are more likely to have lower average socioeconomic status. In turn, math teacher expectations in states that use CRTs are more predictive of math test scores because teachers calibrate (and potentially over-adjust) their expectations for poor students to make them more realistic.

Are accountability standards too high for the average student to meet? When first introduced, accountability systems targeted low-performing students under the guise that introducing achievement goals would create an incentive for students to want to do well, which would in turn raise student achievement. However, increasing evidence suggests that teachers interact differently with high- and low-performing students, such that teachers are more likely to favor better students. If teachers do in fact hold students to different standards, then teachers may respond differently to accountability interventions. In the next section, I revisit the analysis presented in Table 4.3, and examine whether teachers' responses to accountability interventions differ for high- and low-performing students.

Do Teachers Calibrate Their Responses to Accountability Interventions?

As discussed in Chapter 3, some research suggests that teachers' interactions with high- and low-expectancy students differ according to their expectations of students. Research also suggests that teachers' responses to accountability interventions may differ for high- and low-performing students (Hamilton et al. 2007). Table 4.6 examines whether teachers' responses to accountability interventions differs for low- and high-performing students. For this analysis I revisit the analysis presented in Table 4.3 and regress math teacher expectations on the accountability interventions separately for the bottom quartile, bottom half, and top half of students. In particular, I use tenth grade math test scores to identify the three achievement groups from within each race group. I then aggregated the students according to their respective achievement group. As a result, the top and bottom half achievement groups have identical distributions of students by race. I use this within-race group strategy to identify the achievement groups because low-performing students are much more likely to be black or Hispanic, and therefore the results are not solely a function of student race.[10]

Comparing the coefficients for each achievement group, the results suggest that teachers take into account the quality of their students when assessing the likelihood that the students will meet performance expectations. For example, math teacher expectations in early adopter states are slightly higher for the lower

---

[10] Additional analysis show that separating the sample according to the full distribution rather than by race group produces the same results.

performing students than for their higher performing counterparts. By contrast, NRTs are associated with lower math teacher expectations, especially for the higher performing students. It is possible that math teachers lower their expectations for students because NRT scores convey to teachers very little about actual student knowledge. CRTs, on the other hand, reduce teacher expectations for low-performing students, but have no effect on expectations for high-performing students. The even larger coefficient for the bottom quartile of students further suggests that math teachers use test scores to justify their already low expectations of low-performing students. Given that most states administer CRTs, these findings contradict the claim that pressuring teachers to raise their expectations will raise student achievement. Rather, teachers may actually *lower* their expectations, especially for the lowest performing students. Although weaker, exit exams have a similar influence on math teacher expectations insofar as math teacher expectations are slightly lower for the lower performing students only. Finally, neither PATs nor testing frequency have a strong association with math teacher expectations, which is not surprising given that the associations presented in Table 4.3 were already weak.

In sum, further examination of the relationship between accountability interventions and math teacher expectations suggests that teachers take into account student performance when responding to accountability expectations. In other words, teachers adjust their expectations for students differently by student performance. For example, recall from Table 4.3 that math teacher expectations are slightly lower in states that use CRTs. The findings presented in Table 4.6 show that this negative association is a function of teachers calibrating their expectations such that they are more realistic, or less hopeful, of student performance. In other words, pressuring teachers to raise their expectations for low-performing students is difficult because teachers weigh evidence differently across student-performance levels.

Table 4.6: Accountability Intervention Coefficients from Regression Models Predicting Math Teacher Expectations, Separately for the Bottom Quartile, Bottom Half, and Top Half of Students

	Bottom Quartile	Bottom Half	Top Half
Early Implementer	.19	.07	-.06
	(.10)	(.08)	(.08)
	Bottom Quartile	Bottom Half	Top Half
PAT	.05	-.04	.01
	(.10)	(.08)	(.07)
	Bottom Quartile	Bottom Half	Top Half
NRT	-.09	-.03	-.19
	(.09)	(.07)	(.07)
	Bottom Quartile	Bottom Half	Top Half
CRT	-.44	-.35	.00
	(.25)	(.16)	(.16)
	Bottom Quartile	Bottom Half	Top Half
Exit Exam	-.05	-.11	.01
	(.09)	(.07)	(.07)
	Bottom Quartile	Bottom Half	Top Half
Testing Index	-.03	-.04	-.07
	(.05)	(.04)	(.03)
Student Characteristics	✓	✓	✓
Teacher Characteristics	✓	✓	✓
Ubanicity, School Enrollment, and Number of Full-Time Teachers	✓	✓	✓

Sources: State-Level Accountability Characteristics and the Education Longitudinal Study (2002 wave)
Notes: See Table 4.3 notes. N = 2,231 students enrolled in 507 schools for the bottom quartile, N = 4,455 students enrolled in 548 schools for the bottom half, and N = 4,490 students enrolled in 549 schools for the top half of the distribution.

DISCUSSION

Scholarship on accountability testing is still in its infancy. The extant research primarily examines the association between accountability system characteristics and student achievement, attributing changes to accountability interventions. However, it is not easy to make causal claims because it is difficult to identify causal mechanisms (Schiller and Muller 2003). Too often, research ignores the social processes between state-level achievement expectations and student achievement. In this chapter I examined the role of teachers, first by examining how they respond to accountability interventions, and second by examining if teachers communicate accountability expectations to students. Rather than treat all accountability systems as interchangeable, I also examined how teacher responses vary by characteristics of accountability systems.

The results show that math teachers respond adversely to accountability interventions, at least in adjusting their expectations of students. They tend to decrease, rather than increase, their expectations of students, especially the low-performing students. Overall, I find some evidence of an accountability effect on student achievement, but the effect tends to be small. I also find some evidence that teachers communicate accountability expectations to students, but that teachers appear to calibrate their expectations differently for low- and high-performing students. Thus, the findings suggest that teachers mediate accountability expectations, and, as a result, not all students benefit equally from accountability interventions.

These findings may partly explain why recent attempts to increase accountability expectations have failed to substantially narrow achievement gaps. If anything, increasing accountability testing provides teachers with additional information to justify their already low expectations or lower their expectations even more. For the lowest performing students, teachers' even lower expectations may translate to a harsh reality insofar as students accept their teachers' low expectations as their own. Thus, it is not necessarily the "soft bigotry of low expectations" that policymakers should be concerned about, but rather the "hard bigotry of low expectations" or the "harsh reality of diminished expectations." In the next chapter, I revisit the analyses from Chapters 3 and 4, focusing on how student race determines teacher expectations, which in turn affects the race achievement gap.

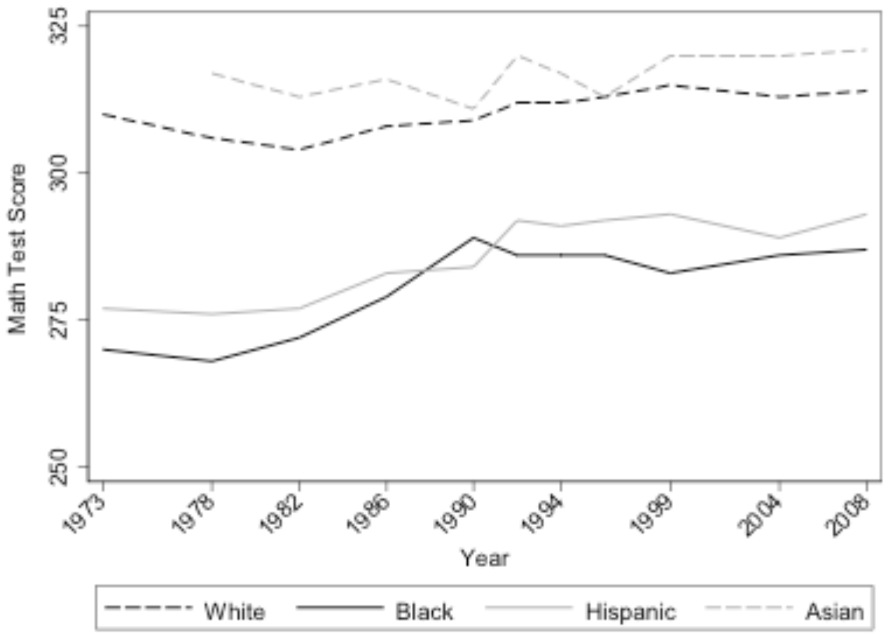
CHAPTER 5

DO TEACHERS' RESPONSES TO ACCOUNTABILITY INTERVENTIONS CONTRIBUTE TO ACHIEVEMENT GAPS BETWEEN WHITE, BLACK, AND HISPANIC STUDENTS?

In Chapters 3 and 4, I examined the association between teacher expectations and student achievement and then how accountability interventions impact this relationship. The findings confirm that teacher expectations are strong predictors of student achievement, but offer little evidence that teachers respond positively to state pressure to raise their expectations for students. In fact, the analysis presented in Chapter 4 suggests that rather than raising their expectations of students, teachers appear to use the information gathered from tests to lower their expectations or to justify their already low expectations of students.

One of the primary goals of policymakers is to close the achievement gaps between white, black, and Hispanic students. Figure 5.1 presents math and reading test score trends by race for students aged 17 years from the mid-1970s to 2008 (NCES 2010). It shows that white and Asian students continue to outperform black and Hispanic students in both math and reading. Whereas average test scores for white and Asians students have remained relatively constant over time, test scores for black and Hispanic students increased slightly between the 1980s and 1990s, but have remained relatively constant over the past two decades. A closer examination reveals that the decline in the black-white achievement gap was due primarily to greater test score gains among the lowest performing black students, which Hedges and Nowell (1998) argue can be explained by increases in socioeconomic status and educational

(a)



(b)

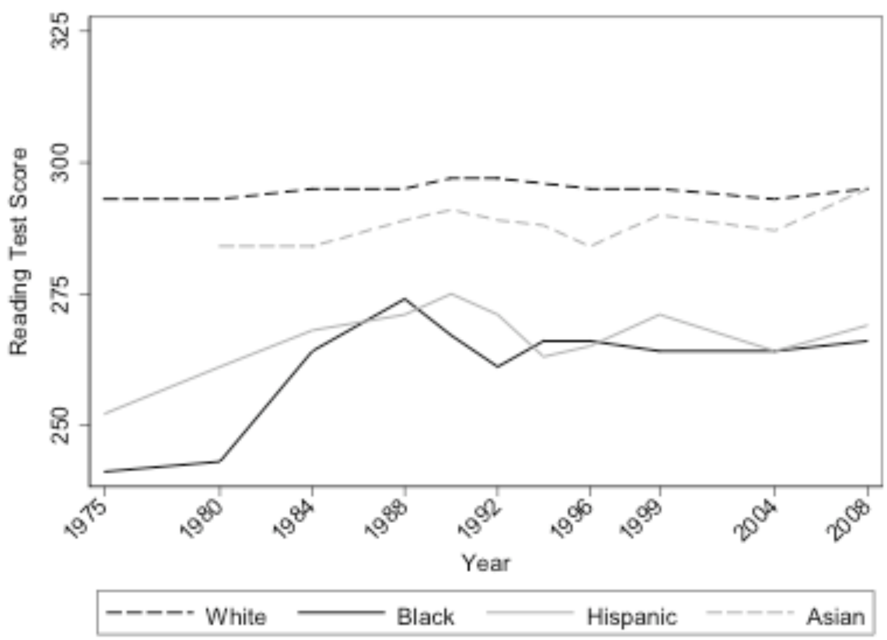


Figure 5.1: NAEP Math and Reading Test Score Trends for Students Aged 17 Years, by Race-Ethnicity
Source: National Center for Education Statistics (2010)

expectations. However, despite their achievement gains, black children start elementary school one year behind their white peers in skills. By the time they finish high school, black students trail white students by four years in skills (Phillips, Crouse, and Ralph 1998). Even if they start elementary school with the same math, reading, and vocabulary skills, the achievement gap widens over time such that at high school completion, nearly one half of the black-white test score gap can be attributed to the initial gap when students start school (Phillips et al. 1998). The race gaps in test scores are visible in other education outcomes. For example, black and Hispanic students are much more likely to be retained and drop out of school than their white counterparts (NCES 2010).

Scholars have offered many explanations for the overall lower performance of black and Hispanic students, including biological differences, neighborhood and school effects, resource deprivation, differences in learning opportunities, and adopting an oppositional culture. Recently, they have directed more attention to the student-teacher relationship, arguing that a student's race affects how teachers perceive and interact with students. In the remainder of this chapter I discuss the literature on student race and the student-teacher relationship. I then revisit the findings presented in Chapters 3 and 4 and examine how student race determines teacher expectations and whether the ways in which teachers mediate accountability expectations differs by students race.

STUDENT RACE AND THE STUDENT-TEACHER RELATIONSHIP

Racial achievement gaps have created a legacy of expectations whereby teachers associate student race with certain behaviors and achievement expectations that may not be associated with actual performance and ability (Correll and Ridgeway 2003). Figlio (2005) shows that teachers form achievement expectations from as little information as a student's race-typed name. The literature on the student-teacher relationship offers abundant evidence that teachers are more likely to perceive black and Hispanic students as being more disruptive and less academically oriented, even if they exhibit more pro-school behavior (Ainsworth-Darnell and Downey 1998; Downey and Pribesh 2004; Farkas, Grobe, Sheehan, and Shuan 1990; Ferguson 2000; Ferguson 1998; Porter, Floden, Freeman, Schmidt, and Schwille 1988). Teachers also tend to have less confidence in black students' ability to learn difficult material (Prime and Miranda 2006). Teachers' own race may affect their perceptions of black and Hispanic students (Downey and Pribesh 2004; Ehrenberg, Goldhaber, and Brewer 1985), and white teachers' more negative perceptions of black students may be evidence of white teacher bias rather than students' oppositional culture (Downey and Pribesh 2004).

Teachers' perceptions of student ability and their expectations affect how they interact with students (Ferguson 2003; Porter et al. 1988; Rubovits and Maehr 1973; Tenenbaum and Ruck 2007) and the amount of effort they put into helping their students (Diamond, Randolph, and Spillane 2004; Prime and Miranda 2006). In their meta-analysis of teacher expectations, Tenenbaum and Ruck (2007) find evidence that teachers offer more positive feedback to white students, on average, than to their black and Hispanic peers. In their analysis of urban elementary schools, Diamond and colleagues (2004) argue that teacher expectations and their sense of responsibility for student learning are related to the student composition. They find evidence that teachers in schools with greater proportions of African American children (to white children) tend to have lower overall expectations of students, African American and white, and to feel less responsible for student learning. They are also more likely to attribute the poor performance of their low-income, black students to factors unrelated to learning, such as students' lives outside of school.

For example, Ann Ferguson's (2000) *Bad Boys: Public Schools in the Making of Black Masculinity* describes how teachers in an urban school near Los Angeles go so far as to single out black children and treat them as adults. In her book, Ferguson describes how teachers openly admit to believing that their students are destined for lives of crime and poverty and treating them as such. Students, even the high achievers, responded by assuming the role their teachers defined for them, thus confirming their teachers' beliefs and perpetuating a legacy of race-based expectations. Teachers who behave in similar ways as those described in Ferguson's book actually *create* a self-fulfilling prophecy, which brings to question whether accountability interventions effectively pressure teachers to adjust their expectations for students. The remainder of this chapter examines how student race affects the student-teacher relationship and how teachers respond to accountability pressure.

RESULTS

In the remainder of this chapter I examine how student race affects (1) teacher expectations of students' future educational attainment, (2) teachers' responses to accountability interventions, (3) the association between teacher expectations and student achievement, and (4) if teachers communicate accountability expectations to students. I present only the main findings, focusing the discussion on four race/ethnic groups: white, black, Hispanic, and Asian students. [See Appendix D for additional analysis as well as results for Native American and multiracial students.]

Before examining how student race influences teacher expectations, it is worth considering how students differ by race/ethnicity. Table 5.1 presents the means and standard deviations for the primary outcome and explanatory variables separately for white, black, Hispanic, and Asian students. Consider first race differences in math test scores. Recall from Table 3.1 in Chapter 3 that the full analysis analytic sample (N =

8,945) has an average tenth grade math test score of 41.96 points, with a standard deviation of 14.14. Table 5.1 shows that this average result masks substantial race differences in math test scores. White and Asian students scored an estimated average of 4.03 and 5.71 points above the sample mean, or .29 and .40 standard deviations respectively. In contrast, black and Hispanic students scored an estimated average of 8.83 and 6.91 points below the sample mean, or .62 and .49 standard deviations respectively. The mean twelfth grade math test scores and test score gains follow a similar pattern, where Asian students have the highest average math test scores, followed by white students, then Hispanic and black students.

Math teacher expectations of students also differ substantially by student race. On average, teachers expect black and Hispanic students to complete an additional two years of schooling beyond high school, or 14.06 and 14.12 years respectively. Math teachers expect white and Asian students, by contrast, to complete an average of three to four years of post-secondary schooling, or 15.05 and 15.86 years respectively. Assuming that math teacher expectations are accurate, predictions of expected years of schooling, and that students meet but do not exceed their teacher's expectations, this race difference translates into the difference between obtaining an Associate's degree versus a Bachelor's degree. Considering the returns to education, the difference between a Bachelor's degree and Associate's degree has potentially serious implications for racial inequality (Kane and Rouse 1995).

The remainder of Table 5.1 presents evidence of race-based differences on other dimensions related to educational achievement. For example, Hispanic and black students have slightly lower socioeconomic status. Black students are the least likely to live in two-parent families and are the most likely to have been held back a grade and/or to have had a behavior problem in the past. Hispanic students are the least likely to be enrolled in the academic track (i.e., 41 percent), while nearly 50

Table 5.1: Means and Standard Deviations for the Primary Variables of Interest, Separately for White, Black, Hispanic, and Asian Students

	White		Black	
	Mean	S.D.	Mean	S.D.
Achievement Test Scores				
IRT estimated number right (10th grade)	45.99	13.35	33.13	11.06
IRT estimated number right (12th grade)	50.99	14.53	37.04	12.32
Gain Score (12th – 10th grade right)	4.52	6.54	3.97	5.26
Math Teacher Expectations (in years)	15.05	2.00	14.06	2.12
Female	.50		.50	
Family Background				
Mother's education (in years)	13.81	2.21	13.37	2.28
Father's education (in years)	13.94	2.50	13.33	2.49
SEI score of mother's occupation in 2002 (GSS 1989 coding)	46.93	13.01	42.97	12.29
SEI score of father's occupation in 2002 (GSS 1989 coding)	45.84	11.73	41.67	10.96
Family income (natural log)	10.86	.84	10.11	1.42
Two-parent family	.80		.50	
Academic Past (as reported by parent)				
Learning disability	.13		.13	
Ever in remedial math	.09		.12	
Ever held back	.10		.23	
Ever have behavior problem	.07		.14	
Academic Track (Academic is the reference category)				
General	.40		.34	
Vocational	.09		.17	
Math Teacher Characteristics				
Female	.57		.56	
Black	.02		.21	
Hispanic	.02		.05	
Asian	.01		.04	
Native American	.01		.00	
Multiracial	.01		.02	
Age (in years)	42.42	11.14	43.47	11.14
Certified degree or higher	.54		.51	
Teaching experience (in years)	14.92	10.42	15.58	10.88
Urbanicity (Suburban is the reference category)				
Urban	.17		.48	
Rural	.28		.11	
Region (Midwest is the reference category)				
Northeast	.19		.17	
South	.32		.58	
West	.18		.07	
School Characteristics				
School enrollment	1304.25	754.32	1475.46	724.46
Full-time teachers	75.95	40.46	85.47	37.99
Percent minority	20.14	19.74	61.57	27.60
Percent free-lunch	16.05	12.78	37.33	21.29
Percent college prep	57.92	28.75	49.47	28.03
Percent general track	48.66	35.44	49.94	30.20
Percent vocational track	17.41	19.91	21.70	22.14
Percent remedial math	6.45	8.71	7.04	10.08
9th Grade GPA	2.84	.77	2.21	.72

	Hispanic		Asian	
	Mean	S.D.	Mean	S.D.
Achievement Test Scores				
IRT estimated number right (10th grade)	35.05	12.54	47.67	15.75
IRT estimated number right (12th grade)	39.87	13.77	53.98	16.90
Gain Score (12th – 10th grade)	4.47	6.38	5.12	8.10
Math Teacher Expectations (in years)	14.12	2.06	15.86	2.15
Female	.51		.50	
Family Background				
Mother's education (in years)	12.17	2.24	13.57	2.73
Father's education (in years)	12.34	2.45	14.24	3.06
SEI score of mother's occupation in 2002	40.01	11.40	43.59	12.10
SEI score of father's occupation in 2002	40.19	10.44	45.00	12.47
Family income (natural log)	10.19	1.38	10.33	1.53
Two-parent family	.76		.84	
Academic Past (as reported by parent)				
Learning disability	.11		.07	
Ever in remedial math	.10		.08	
Ever held back	.16		.09	
Ever have behavior problem	.09		.05	
Academic Track				
General	.44		.31	
Vocational	.15		.11	
Math Teacher Characteristics				
Female	.50		.56	
Black	.06		.04	
Hispanic	.19		.05	
Asian	.04		.13	
Native American	.02		.01	
Multiracial	.03		.04	
Age (in years)	42.39	11.21	42.68	11.00
Certified degree or higher	.54		.57	
Teaching experience (in years)	13.06	10.28	14.43	10.33
Urbanicity				
Urban	.46		.43	
Rural	.09		.09	
Region				
Northeast	.14		.19	
South	.29		.18	
West	.45		.47	
School Characteristics				
School enrollment	1985.50	962.50	1970.96	879.27
Full-time teachers	104.15	47.00	100.86	41.16
Percent minority	63.01	28.14	53.32	29.15
Percent free-lunch	36.08	21.75	22.82	18.02
Percent college prep	50.73	30.59	59.43	31.24
Percent general track	58.42	20.91	58.11	36.36
Percent vocational track	18.37	20.91	14.63	18.56
Percent remedial math	8.25	9.50	8.23	12.28
9th Grade GPA	2.40	.83	2.97	.79

Source: Education Longitudinal Study (2002 and 2004 waves)
Notes: See Table 3.1 notes. For all base-year wave variables, there are 5,104 white students, 1,201 black students, 1,256 Hispanic students, and 875 Asian students. For the twelfth grade test scores and math test score gains, the sample was restricted to respondents who were enrolled in the same school and in the twelfth grade. N = 4,056 white students, 800 black students, 864 Hispanic students and 695 Asian students.

percent of both white and black students and 60 percent of Asian students are enrolled in the academic track.

These differences extend beyond student characteristics and to the types of schools that they attend. In particular, black, Hispanic, and Asian students are more likely to attend urban schools that are more likely have high enrollment rates, disproportionate numbers of minority students, and high poverty rates. Asian and Hispanic students are concentrated in the West, while black students are more likely to live in the South. White students, on the other hand, are more likely to attend predominantly white, suburban schools in the Midwest and South. Next, I examine whether teachers take into account these differences when constructing their expectations of students' future educational attainment. In other words, conditional on student race, do teachers offer more (or less) weight to particular characteristics?

Does Student Race Influence Teacher Expectations of Students' Future Educational Attainment?

To address this question, I regress math teacher expectations on the student, teacher, and school predictor variables separately by student race. The models presented in Table 5.2 are analogous to those presented in Table 3.2 in Chapter 3, where Model 1 includes student characteristics only, Model 2 adds teacher characteristics, and Model 3 adds school characteristics. Models 3b and 4 are presented as additional tests of the Model 3 findings. Model 3b tests the robustness of the Model 3 findings by adjusting for school-level fixed-effects. Model 4 adds ninth grade GPA as a measure of prior academic performance. I estimate the models separately for each race group in order to examine whether teachers use different criteria to construct their expectations.

I turn first to the Model 1 coefficients, a model in which math teacher expectations are regressed on student characteristics only. The coefficients show that teachers hold students from different race groups to slightly different standards. For example, compared the their male counterparts, math teacher expectations are slightly higher for female students. This difference is largest among Asian students, where females are expected to complete .41 years of schooling more. Teachers do not seem to weigh mother's and father's education and occupational prestige differently across race groups, but there are slight differences in their sensitivity to students' family income and to students. In particular, teachers respond more positively to affluent white and black students than to affluent Asian or Hispanic students. The family structure coefficients indicate that teacher expectations are slightly higher for white students who come from a two-parent family. Although subtle, these differences show evidence that teachers consider students' family backgrounds when considering how far in school they expect students to go.

The largest disparities between race groups pertain to students' academic past, which suggests that teachers are more (or less) forgiving of students with troubled academic histories. For example, teachers penalize white and Asian students who have a learning disability much more than their black and Hispanic counterparts, in part because white and Asian students have more to lose insofar as they tend to have higher test scores than black and Hispanic students. Having been enrolled in remedial math in the past has no influence on teacher expectations for Asian students, but is associated with a .5-year decrease in teacher expectations for white, black, and Hispanic students. Hispanic and white students with a behavior problem are expected

Table 5.2: Coefficients from Regression Models Predicting Math Teacher Expectations, Separately for White, Black, Hispanic, and Asian Students

	Model 1				Model 2			
	White	Black	Hispanic	Asian	White	Black	Hispanic	Asian
Female	.19	.17	.24	.41	18	14	.21	.42
	(.05)	(.13)	(.11)	(.20)	(.05)	(.13)	(.11)	(.18)
Family Background								
Mother's education	.09	.11	.08	.07	.09	.11	.09	.05
(in years)	(.02)	(.04)	(.04)	(.04)	(.02)	(.04)	(.04)	(.04)
Father's education (in	.12	.06	.06	.07	.12	.07	.06	.06
years)	(.01)	(.03)	(.03)	(.04)	(.01)	(.03)	(.03)	(.04)
SEI score of mother's	.01	.00	.01	.01	.01	.00	.01	.01
occupation in 2002	(.00)	(.01)	(.01)	(.01)	(.00)	(.01)	(.01)	(.01)
SEI score of father's	.01	-.00	-.00	.02	.01	-.00	-.00	.02
occupation in 2002	(.00)	(.01)	(.01)	(.01)	(.00)	(.01)	(.01)	(.01)
Family income	.14	.14	.05	.03	.14	.14	.07	.02
(natural log)	(.04)	(.05)	(.06)	(.07)	(.04)	(.05)	(.06)	(.07)
Two-parent family	.15	-.05	.09	.08	.16	.02	.07	.13
	(.08)	(.14)	(.17)	(.20)	(.08)	(.14)	(.18)	(.20)
Past History								
Learning disability	-1.06	-.46	-.60	-1.36	-1.07	-.49	-.61	-1.12
	(.09)	(.24)	(.18)	(.39)	(.09)	(.23)	(.18)	(.41)
Ever in remedial	-.45	-.64	-.55	-.05	-.44	-.62	-.56	-.14
math	(.10)	(.21)	(.18)	(.32)	(.10)	(.20)	(.18)	(.32)
Ever held back	-.77	-.89	-.85	-.50	-.78	-.94	-.77	-.45
	(.10)	(.17)	(.20)	(.62)	(.10)	(.17)	(.19)	(.54)
Ever have behavior	-.82	-.56	-1.07	-.54	-.80	-.59	-1.05	-.55
problem	(.13)	(.21)	(.21)	(.51)	(.13)	(.20)	(.21)	(.52)
Academic Track								
General	-.83	-.45	-.68	-.62	-.80	-.46	-.66	-.57
	(.06)	(.16)	(.12)	(.18)	(.13)	(.15)	(.18)	(.18)
Vocational	-.76	-.74	-.66	-.48	-.76	-.73	-.67	-.54
	(.10)	(.19)	(.18)	(.26)	(.10)	(.19)	(.19)	(.26)
Teacher Characteristics								
Female					.16	.15	.37	.20
					(.06)	(.15)	(.37)	(.18)
Black					.41	.62	.36	.12
					(.20)	(.16)	(.27)	(.46)
Hispanic					.54	.78	.55	-.43
					(.26)	(.57)	(.23)	(.56)
Asian					.49	.75	.28	-.67
					(.24)	(.37)	(.38)	(.27)
Native American					-.37	-1.18	-.81	-.50
					(.58)	(.32)	(.61)	(.62)
Multiracial					.02	.15	.41	.21
					(.23)	(.49)	(.38)	(.51)
Age (in years)					-.01	-.03	-.00	-.04
					(.01)	(.01)	(.01)	(.02)
Certified degree or					.10	.30	.15	-.20
higher					(.07)	(.14)	(.16)	(.18)
Teaching experience					.01	.01	-.00	.05
(in years)					(.01)	(.01)	(.01)	(.01)
Constant	15.56	14.70	14.68	16.07	15.38	14.26	14.29	16.14
R^2	.28	.16	.16	.18	.29	.20	.18	.22

Table 5.2: Continued

	Model 3				Model 3b			
	White	Black	Hispanic	Asian	White	Black	Hispanic	Asian
Female	.18	.14	.19	.40	.17	.22	.16	.22
	(.05)	(.13)	(.11)	(.18)	(.06)	(.15)	(.15)	(.19)
Family Background								
Mother's education (in years)	.09	.10	.09	.03	.07	.08	.08	.03
	(.02)	(.03)	(.04)	(.04)	(.02)	(.04)	(.04)	(.05)
Father's education (in years)	.12	.06	.06	.07	.11	.04	.01	.03
	(.01)	(.03)	(.03)	(.04)	(.02)	(.04)	(.03)	(.05)
SEI score of mother's occupation in 2002	.01	.00	.01	.01	.01	-.00	.00	-.00
	(.00)	(.01)	(.01)	(.01)	(.00)	(.01)	(.01)	(.01)
SEI score of father's occupation in 2002	.00	-.00	-.00	.02	.00	-.00	-.00	.03
	(.00)	(.01)	(.01)	(.01)	(.00)	(.01)	(.01)	(.01)
Family income (natural log)	.10	.13	.07	.04	.10	.13	.13	-.06
	(.04)	(.05)	(.07)	(.07)	(.04)	(.07)	(.09)	(.08)
Two-parent family	.17	.03	.08	.24	.21	.18	-.03	.00
	(.08)	(.14)	(.18)	(.20)	(.08)	(.16)	(.20)	(.25)
Past History								
Learning disability	-1.09	-.50	-.58	-.98	-1.10	-.57	-.57	-.87
	(.08)	(.22)	(.18)	(.41)	(.09)	(.29)	(.27)	(.47)
Ever in remedial math	-.42	-.64	-.58	.01	-.39	-.38	-.62	.12
	(.10)	(.20)	(.18)	(.30)	(.10)	(.26)	(.24)	(.40)
Ever held back	-.80	-.94	-.82	-.28	-.75	-.81	-.67	-.61
	(.10)	(.16)	(.19)	(.44)	(.10)	(.21)	(.21)	(.46)
Ever have behavior problem	-.79	-.60	-1.06	-.71	-.85	-.71	-1.22	-.79
	(.13)	(.20)	(.20)	(.53)	(.12)	(.29)	(.25)	(.19)
Academic Track								
General	-.76	-.42	-.65	-.51	-.81	-.47	-.68	-.79
	(.06)	(.15)	(.13)	(.17)	(.06)	(.17)	(.16)	(.19)
Vocational	-.71	-.71	-.61	-.47	-.66	-.79	-.81	-.36
	(.10)	(.19)	(.19)	(.27)	(.10)	(.19)	(.23)	(.32)
Teacher Characteristics								
Female	.12	.15	.32	.15	.09	.06	.31	.22
	(.06)	(.15)	(.17)	(.18)	(.07)	(.17)	(.19)	(.23)
Black	.31	.53	.24	-.12	.40	.70	1.01	-.54
	(.21)	(.18)	(.31)	(.48)	(.21)	(.25)	(.42)	(.55)
Hispanic	.48	.77	.46	-.49	.31	.63	.54	-.54
	(.23)	(.43)	(.23)	(.49)	(.30)	(.53)	(.27)	(.82)
Asian	.45	.69	.42	-.79	.41	1.54	.59	-.48
	(.25)	(.39)	(.39)	(.34)	(.27)	(.59)	(.49)	(.41)
Native American	-.30	-1.02	-.76	-.82	.39	-971.34	-3.34	-1.23
	(.47)	(.46)	(.50)	(.63)	(.45)	(1962.08)	(.99)	(1.27)
Multiracial	-.06	.10	.37	.24	.24	.17	.94	-.19
	(.23)	(.49)	(.39)	(.46)	(.34)	(.52)	(.47)	(.41)
Age (in years)	-.01	-.03	-.00	-.04	-.01	-.04	-.01	-.05
	(.01)	(.01)	(.01)	(.02)	(.01)	(.01)	(.01)	(.02)
Certified degree or higher	.09	.24	.23	-.17	.06	.15	.47	-.14
	(.07)	(.15)	(.16)	(.18)	(.07)	(.20)	(.19)	(.22)

Table 5.2: Continued

Teaching experience (in years)	.00	.01	-.00	.05	.01	.03	.02	.06
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.02)
Urbanicity								
Urban	.21	.13	-.04	.37				
	(.09)	(.17)	(.18)	(.21)				
Rural	-.08	-.08	.03	.20				
	(.07)	(.20)	(.24)	(.29)				
Region								
Northeast	.32	.56	-.37	.45				
	(.09)	(.27)	(.30)	(.26)				
South	-.03	.15	-.09	.61				
	(.08)	(.19)	(.23)	(.30)				
West	-.21	.03	-.41	.54				
	(.11)	(.36)	(.25)	(.29)				
School Composition								
School size	-.00	-.00	-.00	-.00				
	(.00)	(.00)	(.00)	(.00)				
Full-time teachers	.00	.00	.01	.00				
	(.00)	(.01)	(.00)	(.01)				
Percent minority	.00	.01	-.00	.00				
	(.00)	(.01)	(.00)	(.00)				
Percent free lunch	-.00	-.00	.01	-.00				
	(.00)	(.01)	(.01)	(.01)				
Percent college prep	.00	.00	.00	.00				
	(.00)	(.00)	(.00)	(.00)				
Percent general	.00	-.00	-.00	-.00				
	(.00)	(.00)	(.00)	(.00)				
Percent vocational	-.00	-.00	-.01	-.00				
	(.00)	(.00)	(.00)	(.01)				
Percent remedial math	-.00	-.01	.00	-.02				
	(.00)	(.01)	(.01)	(.01)				
9th Grade GPA								
Constant	15.36	14.05	14.56	15.51	15.39	15.43	14.25	16.28
R^2	.31	.21	.20	.25	.40	.48	.49	.56

Table 5.2: Continued

	Model 4			
	White	Black	Hispanic	Asian
Female	-.06	-.02	-.05	-.07
	(.05)	(.12)	(.09)	(.15)
Family Background				
Mother's education (in years)	.04	.07	.07	.06
	(.01)	(.03)	(.03)	(.03)
Father's education (in years)	.07	.03	.02	.06
	(.01)	(.03)	(.02)	(.03)
SEI score of mother's occupation in 2002	.01	-.00	.01	-.00
	(.00)	(.00)	(.01)	(.01)
SEI score of father's occupation in 2002	.00	-.00	.01	.01
	(.00)	(.01)	(.01)	(.01)
Family income (natural log)	.05	.12	.06	.04
	(.04)	(.05)	(.06)	(.06)
Two-parent family	-.00	-.07	-.01	.21
	(.07)	(.13)	(.17)	(.17)
Past History				
Learning disability	-.69	-.42	-.40	-.69
	(.09)	(.21)	(.18)	(.38)
Ever in remedial math	-.23	-.54	-.45	.04
	(.09)	(.18)	(.16)	(.28)
Ever held back	-.40	-.50	-.66	.22
	(.10)	(.16)	(.18)	(.33)
Ever have behavior problem	-.38	-.20	-.39	-.22
	(.14)	(.19)	(.12)	(.46)
Academic Track				
General	-.37	-.15	-.39	-.05
	(.05)	(.14)	(.12)	(.15)
Vocational	-.38	-.45	-.30	-.05
	(.10)	(.18)	(.17)	(.23)
Teacher Characteristics				
Female	.06	.15	.14	.11
	(.05)	(.13)	(.15)	(.14)
Black	.41	.62	.43	.30
	(.18)	(.19)	(.27)	(.35)
Hispanic	.24	.98	.30	-.40
	(.26)	(.48)	(.22)	(.40)
Asian	.40	.49	.09	-.56
	(.23)	(.40)	(.36)	(.23)
Native American	-.43	-.37	-.38	-.01
	(.55)	(.44)	(.48)	(.91)
Multiracial	-.02	.08	.29	.57
	(.21)	(.43)	(.34)	(.34)
Age (in years)	-.01	-.03	-.00	-.02
	(.00)	(.01)	(.15)	(.01)
Certified degree or higher	.03	.16	-.08	-.10
	(.06)	(.14)	(.15)	(.14)
Teaching experience (in years)	-.00	.01	-.01	.02
	(.00)	(.01)	(.01)	(.01)

Table 5.2: Continued

Urbanicity				
Urban	.19	.32	-.08	.25
	(.09)	(.16)	(.7)	(.16)
Rural	-.06	-.15	.20	.32
	(.08)	(.19)	(.22)	(.25)
Region				
Northeast	.49	.59	-.06	.63
	(.09)	(.25)	(.30)	(.24)
South	.08	-.04	-.19	.69
	(.08)	(.18)	(.23)	(.25)
West	-.32	-.04	-.40	.08
	(.11)	(.32)	(.24)	(.23)
School Composition				
School size	-.00	-.00	-.00	.00
	(.00)	(.00)	(.00)	(.00)
Full-time teachers	.00	.00	.01	.00
	(.00)	(.00)	(.00)	(.00)
Percent minority	.01	.01	.00	.01
	(.00)	(.00)	(.00)	(.00)
Percent free lunch	-.01	-.01	.01	-.00
	(.00)	(.01)	(.01)	(.00)
Percent college prep	.00	.00	.00	.01
	(.00)	(.00)	(.00)	(.00)
Percent general	.00	-.00	-.00	.00
	(.00)	(.00)	(.00)	(.00)
Percent vocational	-.00	-.00	-.01	-.00
	(.00)	(.00)	(.00)	(.00)
Percent remedial math	.00	-.00	.00	-.00
	(.00)	(.01)	(.01)	(.01)
9th Grade GPA	1.23	1.22	1.09	1.53
	(.04)	(.10)	(.07)	(.10)
Constant	15.31	13.92	14.67	15.60
R^2	.47	.35	.36	.46

Source: Education Longitudinal Study (2002 wave)
Notes: See Table 5.1 notes.

to complete 1.07 and .82 years of schooling less, versus a decrease of .56 and .54 for black and Asian students, respectively. These findings offer evidence that conditional on student race, teachers may be more inclined to forgive certain behaviors. This result is consistent with other scholarship that shows that teachers are most supportive of high-expectancy students when they show signs of achievement gains, but are more critical of low expectancy students when the students show greater-than-expected achievement gains (Rosenthal and Jacobson 1968). The results also show slight differences in the rate at which teachers adjust their expectations, according to the academic track in which students are enrolled. For example, white students are penalized more for being enrolled in the general track, and Asian students for being enrolled in the vocational track.

Together, the results from Model 1 offer evidence that the criteria on which students are judged differs according to student race, meaning that teacher expectations are biased. More specifically, teachers have pre-existing ideas of students' academic abilities, which they use as a foundation for their expectations. As a result, some students appear to suffer more from the negative consequences of high teacher expectations (e.g., Asian and white students who have a learning disability).

The coefficients for Model 2 indicate that the differences between race groups are not simply a function of students having different teachers. They offer some evidence, although weak, that teachers' own characteristics influence their expectations for students' future educational attainment. For example, white, black, and Hispanic students with non-white teachers (i.e., black, Hispanic, or Asian) are held to slightly higher standards than their counterparts who have white teachers. By contrast, Asian students with non-white teachers are held to lower standards, especially those with a same-race math teacher.[11] Together, the lower expectations for Asian students, in combination with the higher standards for white, black, and Hispanic students, suggests that math teacher expectations are not only sensitive to student race, but affected by racial stereotypes.

To what extend does past performance influence these dynamics? Model 4 addresses this question by including ninth grade GPA as a measure of students' past performance. As in Chapter 3, including prior academic performance in the model reduces the size of the estimated effect of the student predictor variables, especially the past history and track variables. It does not, however, explain the association away entirely. Thus, prior academic performance continues to be an important signal for teachers.

Race Differences in Teacher Expectations: Do Accountability Interventions Make a Difference?

Are teachers responding to accountability interventions? In other words, are accountability interventions pressuring teachers to raise their expectations of traditionally low-performing students? In this section, I offer two analyses that investigate whether teachers' responses to accountability interventions differ by student race. To this end, Table 5.3 presents the coefficients from regression models predicting math teacher expectations. Similar to Table 4.3 in Chapter 4, each type of accountability intervention is modeled separately.

Three models are presented for each type of intervention, all of which adjust for the student, teacher, and school level predictor variables. Model 1 (which

[11] These findings should be interpreted with caution because the proportion of students with non-white teachers is quite small. For example, the results suggest that math teacher expectations are substantially low for black students who have a Native American teacher. Additional analysis (not presented) shows that removing the dummy variable for Native American teachers does not affect the results.

corresponds to Model 3 from Table 4.3 in Chapter 4) includes a dummy variable for each race group and accountability intervention. Model 2 adds an interaction term between each race category and accountability intervention in order to examine how teacher responses to accountability interventions differ by student race. And Model 3 adds prior academic performance to the Model 2 specification. Unlike the analysis presented in Table 5.2, which estimated models separately by student race, the analyses presented in Tables 5.3 and 5.4 employ the full analytic sample (N = 8,945).[12]

Before discussing the findings for each accountability intervention, it is worth considering how adding the accountability intervention measures to the models affects the relationship between race and math teacher expectations. Recall Model 3 from Table 3.2 in Chapter 3, where math teacher expectations were regressed on student, teacher, and school characteristics: these results showed that math teacher expectations for black and Hispanic students were lower than for whites and Asian students. Adding the accountability intervention variables to the model (see Table 5.3 models) does not affect either the direction or the size of the student race coefficients, indicating that student race is an important predictor of teacher expectations, regardless of accountability systems. Models 2 and 3, however, suggest that some of the association between student race and teacher expectations is attributed to teachers' responses to accountability pressure, as indicated by the interaction terms. In other words, accountability interventions increase the strength of the association between student race and math teacher expectations. Race thus influences teacher expectations both directly and indirectly.

Next I consider how teachers respond to pressure to raise their expectations of students. The findings discussed in Chapter 4 (also presented in Model 1 of Table 5.3)

[12] The coefficients for Native American and multiracial students are suppressed in Tables 5.3 and 5.4. See Table C2 in Appendix C for results from models where each race is modeled separately.

show that how math teachers respond to accountability pressure is a function of the type or strength of accountability intervention. Math teachers either do not respond at all or *lower* their expectations of students. Models 2 and 3 show that student race is an important determinant of how math teachers respond to pressure to raise their expectations. CRTs are a typical case. Adjusting for the predictor variables only, CRTs are associated with a .17-year decrease in math teacher expectations (see Model 1). Model 2, however, shows that teachers' responses vary substantially according to student race. The main effect indicates that math teacher expectations are .21-years lower for white students in states that use CRTs. Teachers in states that use CRTs expect black students to complete nearly one year less of schooling (i.e., -.21 - .70 = -.91) than their same-race peers in states that do not use CRTs. By contrast, teachers expect Asian and Hispanic students to complete 1.44 (i.e., -.21 + 1.65 = 1.44) and .28 (i.e., -.21 + .49 = .28) years of schooling more than their same-race peers in states that do not use CRTs. Adjusting for prior academic performance explains away the size of the direct effect, but not the interaction effects. In other words, student race is an important determinant of how teachers respond to accountability pressure. Although the associations for the other accountability characteristics are quite small, they offer additional support for the findings presented in Table 4.3 in Chapter 4, where I found that teachers do not raise their expectations for all students.

The findings presented so far have specifically looked at the importance of student race as a determinant of math teacher expectations. They offer ample evidence that teachers hold black and Hispanic students to lower standards and that pressuring teachers to raise their expectations may in fact have the opposite effect, which may exacerbate the black-white achievement gap. Historically, black and Hispanic students tend to have lower performance levels than white and Asian

Table 5.3: Coefficients for Student Race, Accountability Interventions, and the Interaction Between Student Race and Accountability Interventions from Models Predicting Math Teacher Expectations, Separately by Accountability Intervention

	Accountability Intervention: Early Implementer		
	Model 1	Model 2	Model 3
Race			
Black	-.77	-.87	-.49
	(.09)	(.15)	(.14)
Hispanic	-.56	-.52	-.34
	(.09)	(.19)	(.15)
Asian	.59	.86	.60
	(.11)	(.28)	(.22)
Early Implementer	.01	.02	.12
	(.06)	(.07)	(.07)
Interaction Effect			
Black * Early Implementer		.13	.26
		(.17)	(.16)
Hispanic * Early Implementer		-.05	.00
		(.21)	(.17)
Asian * Early Implementer		-.32	-.20
		(.30)	(.24)
	Accountability Intervention: Performance Assessment Test		
	Model 1	Model 2	Model 3
Race			
Black	-.77	-.69	-.24
	(.09)	(.10)	(.09)
Hispanic	-.56	-.55	-.31
	(.09)	(.10)	(.09)
Asian	.59	.65	.47
	(.11)	(.12)	(.10)
PAT	.04	.10	.10
	(.06)	(.07)	(.06)
Interaction Effect			
Black * PAT		-.33	-.20
		(.17)	(.15)
Hispanic * PAT		.08	-.06
		(.23)	(.20)
Asian * PAT		-.25	-.11
		(.24)	(.22)

Table 5.3: Continued

	Accountability Intervention: Norm-Referenced Test		
	Model 1	Model 2	Model 3
Race			
Black	-.78	-.77	-.31
	(.09)	(.11)	(.11)
Hispanic	-.55	-.67	-.44
	(.09)	(.13)	(.13)
Asian	.59	.54	.41
	(.11)	(.16)	(.13)
NRT	-.14	-.18	-.23
	(.06)	(.06)	(.06)
Interaction Effect			
Black * NRT		-.02	.02
		(.16)	(.15)
Hispanic * NRT		.22	.22
		(.18)	(.17)
Asian * NRT		.11	.08
		(.21)	(.17)

	Accountability Intervention: Criterion-Referenced Test		
	Model 1	Model 2	Model 3
Race			
Black	-.77	-.07	.69
	(.09)	(.44)	(.45)
Hispanic	-.56	-1.04	-.57
	(.09)	(.58)	(.51)
Asian	.59	-1.06	-.32
	(.11)	(.93)	(.39)
CRT	-.17	-.21	.00
	(.12)	(.14)	(.13)
Interaction Effect			
Black * CRT		-.70	-.98
		(.43)	(.45)
Hispanic * CRT		.49	.24
		(.58)	(.51)
Asian * CRT		1.65	.76
		(.93)	(.40)

Table 5.3: Continued

	Accountability Intervention: Exit Exam		
	Model 1	Model 2	Model 3
Race			
Black	-.77	-.78	-.15
	(.09)	(.14)	(.15)
Hispanic	-.56	-.53	-.33
	(.09)	(.16)	(.16)
Asian	.59	.60	.42
	(.11)	(.21)	(.17)
Exit Exam	-.03	-.04	.05
	(.06)	(.06)	(.07)
Interaction Effect			
Black * Exit Exam		.01	-.18
		(.16)	(.17)
Hispanic * Exit Exam		-.03	-.00
		(.19)	(.18)
Asian * Exit Exam		-.02	.03
		(.24)	(.18)

	Accountability Intervention: Testing Index		
	Model 1	Model 2	Model 3
Race			
Black	-.77	-.77	-.29
	(.09)	(.09)	(.08)
Hispanic	-.55	-.57	-.35
	(.09)	(.09)	(.09)
Asian	.60	.59	.44
	(.11)	(.11)	(.09)
Testing Index	-.07	-.10	-.08
	(03)	(.04)	(.04)
Interaction Effect			
Black * Testing Index		.01	-.01
		(.07)	(.07)
Hispanic * Testing Index		.12	.13
		(.09)	(.08)
Asian * Testing Index		.13	.10
		(.09)	(.08)
Student Characteristics	✓	✓	✓
Teacher Characteristics	✓	✓	✓
School Characteristics[a]	✓	✓	✓
9th Grade GPA			✓

Sources: State-Level Accountability Characteristics and the Education Longitudinal Study (2002 wave)
Notes: The data are weighted by the NCES post-stratification weight (BYSTUWT) multiplied by the inverse probability of having non-missing information for the primary explanatory variable, math teacher expectations. $N = 8{,}945$ students enrolled in 561 schools.
[a] Region is not included as an adjustment variable.

students. Are the findings presented so far merely a race phenomenon, or do teachers treat all low-performing students similarly?

Table 5.4 re-estimates Model 2 from Table 5.3 separately for three achievement groups. Similar to the analysis presented in Table 4.5 in Chapter 4, I use tenth grade math test scores to identify the bottom quartile, bottom half, and top half of students within each race group. Look first at the top panel, which presents the results for early implementation. Given the history of testing (i.e., tests were first used to monitor low-income student performance), it is not surprising to see that teacher expectations of low- and high-performing black students and low-performing Hispanic students are higher in states that were early adopters of accountability testing than their same-race peers in later adopter states. The remaining panels suggest that math teachers' lower expectations of black and Hispanic students are not simply a function of student performance. Instead, teachers appear to hold black students to lower standards, despite their performance. Accountability interventions are associated with slightly higher math teacher expectations for Hispanic students, but their expectations are not consistently higher for all students.

Consider the association between CRTs and math teacher expectations for black and Hispanic students. Math teacher expectations for black students in CRT states are consistently low for low- *and* high-performing students, at -1.26 for the bottom quartile (i.e., -.47 - .79 = -1.26), -1.25 for the bottom half (i.e., -.39 - .86 = -1.25), and -1.08 (i.e., -.14 - .94 = -1.08) for the top half of students. By contrast, CRTs are associated with higher math teacher expectations for high-performing Hispanic students, but not their low-performing counterparts, at -.42 for the bottom quartile (i.e., -.47 + .05 = -.42), -.32 (i.e., -.39 + .07 = -.32) for the bottom half, and .87 (i.e., -.14 + 1.01 = .87) for the top half. The coefficients also show that math teachers over-adjust their expectations for low-performing Asian students.

Table 5.4: Coefficients for Student Race, Accountability Interventions, and the Interaction Between Student Race and Accountability Interventions from Models Predicting Math Teacher Expectations, Separately for the Bottom Quartile, Bottom Half, and Top Half of Students

	Accountability Intervention: Early Implementer		
	Bottom Quartile	Bottom Half	Top Half
Race			
Black	-1.12	-.99	-1.04
	(.26)	(.19)	(.21)
Hispanic	-.76	-1.01	-.61
	(.28)	(.22)	(.19)
Asian	.35	.27	1.15
	(.48)	(.44)	(.24)
Early Implementer	.09	.02	-.06
	(.12)	(.08)	(.09)
Interaction Effect			
Black * Early Implementer	.34	.04	.13
	(.30)	(.21)	(.24)
Hispanic * Early Implementer	.08	.15	-.07
	(.30)	(.24)	(.21)
Asian * Early Implementer	-.04	.06	-.31
	(.50)	(.45)	(.28)

	Accountability Intervention: Performance Assessment Test		
	Bottom Quartile	Bottom Half	Top Half
Race			
Black	-.83	-.88	-.85
	(.18)	(.13)	(.12)
Hispanic	-.69	-.84	-.70
	(.16)	(.12)	(.11)
Asian	.31	.33	.93
	(.25)	(.16)	(.14)
PAT	.12	.06	.06
	(.11)	(.09)	(.08)
Interaction Effect			
Black * PAT	-.34	-.39	-.28
	(.31)	(.22)	(.22)
Hispanic * PAT	.10	-.18	.31
	(.38)	(.31)	(.27)
Asian * PAT	.00	-.01	-.18
	(.30)	(.30)	(.27)

Table 5.4: Continued

	Accountability Intervention: Norm-Referenced Test		
	Bottom Quartile	Bottom Half	Top Half
Race			
Black	-.81	-.94	-.91
	(.19)	(.14)	(.14)
Hispanic	-.64	-.94	-.83
	(.24)	(.18)	(.15)
Asian	.25	.33	.99
	(.25)	(.20)	(.17)
NRT	.00	-.01	-.20
	(.10)	(.08)	(.08)
Interaction Effect			
Black * NRT	-25	-.06	-.07
	(.29)	(.21)	(.19)
Hispanic * NRT	-.09	.13	.33
	(.27)	(.22)	(.19)
Asian * NRT	.12	.00	-.11
	(.36)	(.26)	(.23)

	Accountability Intervention: Criterion-Referenced Test		
	Bottom Quartile	Bottom Half	Top Half
Race			
Black	-.11	-.10	.02
	(.41)	(.49)	(.41)
Hispanic	-.74	-.94	-1.67
	(1.66)	(1.13)	(.62)
Asian	-1.66	-2.10	-.27
	(.35)	(.20)	(.20)
CRT	-.47	-.39	-.14
	(.31)	(.17)	(.15)
Interaction Effect			
Black * CRT	-.79	-.86	-.94
	(.39)	(.48)	(.42)
Hispanic * CRT	.05	.07	1.01
	(1.67)	(1.14)	(.63)
Asian * CRT	1.99	2.44	1.17
	(.38)	(.24)	(.23)

Table 5.4: Continued

	Accountability Intervention: Exit Exam		
	Bottom Quartile	Bottom Half	Top Half
Race			
Black	-.72	-.87	-.81
	(.26)	(.20)	(.16)
Hispanic	-.81	-.99	-.52
	(.28)	(.18)	(.18)
Asian	.47	.53	.78
	(.36)	(.30)	(.22)
Exit Exam	-.08	-.13	-.01
	(.10)	(.08)	(.08)
Interaction Effect			
Black * Exit Exam	-.24	-.12	-.14
	(.28)	(.22)	(.20)
Hispanic * Exit Exam	.16	-.00	-.18
	(.30)	(.21)	(.19)
Asian * Exit Exam	-.22	-.29	.16
	(.30)	(.32)	(.27)

	Accountability Intervention: Testing Index		
	Bottom Quartile	Bottom Half	Top Half
Race			
Black	-.91	-.96	-.92
	(.16)	(.11)	(.11)
Hispanic	-.67	-.89	-.70
	(.17)	(.12)	(.11)
Asian	.31	.33	.94
	(.21)	(.14)	(.12)
Testing Index	-.04	-.05	-.10
	(.06)	(.05)	(.04)
Interaction Effect			
Black * Testing Index	.01	-.05	-.00
	(.14)	(.10)	(.09)
Hispanic * Testing Index	-.01	.09	.17
	(.12)	(.11)	(.10)
Asian * Testing Index	.03	.06	-.03
	(.18)	(.11)	(.11)
Student Characteristics	✓	✓	✓
Teacher Characteristics	✓	✓	✓
Urbanicity, School Enrollment, and Number of Full-Time Teachers	✓	✓	✓

Sources: State-Level Accountability Characteristics and the Education Longitudinal Study (2002 wave)
Notes: See Table 4.5 notes. N = 2,231 for the bottom quartile, 4,455 for the bottom half, and 4,490 for the top half of the distribution. N = 2,231 students enrolled in 507 schools for the bottom quartile, N = 4,455 students enrolled in 548 schools for the bottom half, and N = 4,490 students enrolled in 549 schools for the top half of the distribution.

So far, I have examined how student race influences the criteria upon which teachers form their expectations and teachers' responses to accountability interventions. The findings from these analyses demonstrate that student race plays an important role in the student-teacher relationship. Findings from the first analysis offer evidence that student race affects how teachers perceive students. Teacher expectations are particularly sensitive to students with academic pasts that indicate either learning or behavior problems. Similar to the results presented in Chapter 4, the findings presented Tables 5.3 and 5.4 show that teachers do not respond positively to accountability interventions: the race differences are persistent no matter the type or strength of accountability system. That is, math teachers do not raise their expectations for the students that are more likely to be held to lower standards. The implications of these findings for policy are discussed in the concluding chapter. In the remainder of this chapter, I examine how student race affects the association between math teacher expectations and student achievement.

Does Student Race Affect the Association Between Teacher Expectations and Student Achievement?

How do racial differences in math teacher expectations affect student performance? Table 5.5 presents the coefficients for math teacher expectations separately by student race for three outcome variables: tenth grade math test, twelfth grade math test, and math test score gains (i.e., 12^{th} – 10^{th} grade). The models presented in each panel correspond to the models presented in Table 3.4 from Chapter 3, where the top row of each panel presents the bivariate association between math teacher expectations and achievement. Models 1 through 4 include the same adjustment variables, as presented in Table 5.2.

First, I examine the association between student race, teacher expectations, and tenth grade math test scores. The coefficients from a naïve model, with no predictor variables, shows that teachers' influence on student achievement varies substantially by student race. Each additional year of schooling that teachers expect students to complete is associated with an increase of 3.94 test score points for white students, 2.54 points for black students, 3.07 points for Hispanic students, and 4.53 points for Asian students. The association for each race group decreases once student, teacher, and school characteristics are taken into account, but the race differences persist (see Models 1, 2, and 3). The coefficients for Model 3 indicate that each additional year of expected schooling is associated with an increase in tenth grade math test scores of 2.93, 1.92, 2.65, and 3.59 points for white, black, Hispanic, and Asian students, respectively. These differences persist even after adjusting for school fixed-effects (see Model 3b). Adding prior academic performance reduces the sizes of the coefficients, especially for white and Asian students (see Model 4). Each additional year of schooling math teachers expect students to complete is associated with an increase of 2.15 and 2.73 test score points for white and Asian students respectively, compared to 1.88 and 2.03 points for black and Hispanic students respectively.

The results for twelfth grade math test scores presented in the middle panel follow a similar pattern: Asian students benefit the most from teacher expectations, followed by white, Hispanic, and then black students. For math test score gains, however, the results tell a slightly different story. Unlike the tenth and twelfth grade findings, where black students benefitted the least from teacher expectations, the math gains models suggest that Hispanic students gain the least between the tenth and twelfth grades. In fact, adjusting for prior academic achievement, the teacher expectations effect disappears entirely for Hispanic students, but not for white, black, or Asian students (see Model 4). In other words, Hispanic students do not make any

Table 5.5: Math Teacher Expectations Coefficients from Regression Models Predicting 10th and 12th Grade Math Tests and Math Test Gains, Separately for White, Black, Hispanic, and Asian Students

	Outcome Variable: 10th Grade Math Test			
	White	Black	Hispanic	Asian
Unadjusted Model: Math teacher expectations	3.94	2.54	3.07	4.53
	(.09)	(.16)	(.20)	(.25)
Model 1: Unadjusted model + student characteristics	2.91	2.10	2.55	3.95
	(.10)	(.15)	(.19)	(.26)
Model 2: Model 1 + teacher characteristics	2.93	2.27	2.63	3.93
	(.11)	(.15)	(.19)	(.25)
Model 3: Model 2 + school characteristics	2.93	1.92	2.65	3.59
	(.10)	(.18)	(.23)	(.30)
Model 3b: Model 2 + school fixed effects	3.06	1.92	2.65	3.59
	(.10)	(.18)	(.23)	(.30)
Model 4: Model 3 + 9th grade GPA	2.15	1.88	2.03	2.73
	(.11)	(.15)	(.20)	(.26)
	Outcome Variable: 12th Grade Math Test			
	White	Black	Hispanic	Asian
Unadjusted Model: Math teacher expectations	4.65	3.15	3.66	5.03
	(.13)	(.23)	(.31)	(.35)
Model 1: Unadjusted model + student characteristics	3.52	2.47	2.89	4.30
	(.14)	(.21)	(.28)	(.34)
Model 2: Model 1 + teacher characteristics	3.55	2.58	3.02	4.36
	(.13)	(.22)	(.27)	(.35)
Model 3: Model 2 + school characteristics	3.55	2.56	3.05	4.13
	(.13)	(.21)	(.26)	(.37)
Model 3b: Model 2 + school fixed effects	3.74	2.02	3.21	4.22
	(.12)	(.28)	(.35)	(.46)
Model 4: Model 3 + 9th grade GPA	2.55	2.25	2.33	2.78
	(.14)	(.22)	(.28)	(.35)
	Outcome Variable: Math Gain (12th – 10th Grade Math Test)			
	White	Black	Hispanic	Asian
Unadjusted Model: Math teacher expectations	.67	.46	.39	.67
	(.08)	(.11)	(.13)	(.14)
Model 1: Unadjusted model + student characteristics	.59	.35	.27	.49
	(.09)	(.12)	(.14)	(.21)
Model 2: Model 1 + teacher characteristics	.59	.34	.24	.52
	(.08)	(.12)	(.13)	(.23)
Model 3: Model 2 + school characteristics	.59	.32	.21	.53
	(.08)	(.11)	(.13)	(.22)
Model 3b: Model 2 + school fixed effects	.59	.39	.18	.59
	(.08)	(.15)	(.17)	(.29)
Model 4: Model 3 + 9th grade GPA	.37	.22	.01	.34
	(.10)	(.12)	(.14)	(.25)

Source: Education Longitudinal Study (2002 and 2004 waves)
Notes: See Table 5.1 notes.

substantial achievement gains between the tenth and twelfth grades. The coefficients for white and Asian students are .37 and .34, respectively, and .22 for black students.

The findings presented in Table 5.5 demonstrate that teacher expectations may be biased, which contributes to the race achievement gap. The achievement gains associated with teacher expectations are smaller for black and Hispanic students than for white and Asian students. Given that black and Hispanic students have lower test scores to begin with, the smaller return to teacher expectations is particularly problematic for the race achievement gap. This finding is, of course, crucial to whether the policy objectives of NCLB are met. Next, I offer two final pieces of analysis that examine how the association between accountability interventions, math teacher expectations and student achievement differs by student race.

Do Teachers' Responses to Accountability Interventions Contribute to Achievement Gaps Between White, Black, and Hispanic Students?

To answer this question, I revisit the analyses presented in Tables 4.4 and 4.5 in Chapter 4 to examine whether the association between accountability interventions, math teacher expectations, and tenth grade math test scores differs by student race. Table 5.6 shows results from models regressing tenth grade math test scores on accountability interventions, math teacher expectations, and the adjustment variables, separately for each race group. Table 5.7 shows the results for a model that allows math teacher expectations to interact with accountability interventions. The top panel of each table presents the math teacher expectations from models that do not include any accountability intervention variables.

First consider how including the accountability intervention variables in the models affects the association between math teacher expectations and tenth grade math test scores. Consistent with the findings presented in Table 4.4, adding the

Table 5.6: Coefficients for Math Teacher Expectations and Accountability Interventions from a Regression Model Predicting 10th Grade Math Test, Separately for White, Black, Hispanic, and Asian Students

	White	Black	Hispanic	Asian
Math Teacher Expectations	2.93	2.26	2.63	3.81
	(.10)	(.14)	(.18)	(.24)

	Accountability Intervention: Early Implementer			
	White	Black	Hispanic	Asian
Math Teacher Expectations	2.93	2.25	2.63	3.80
	(.10)	(.14)	(.18)	(.24)
Early Implementer	-.02	1.01	.06	-.46
	(.49)	(.90)	(.93)	(1.55)

	Accountability Intervention: Performance Assessment Test			
	White	Black	Hispanic	Asian
Math Teacher Expectations	2.93	2.26	2.63	3.80
	(.10)	(.15)	(.18)	(.25)
PAT	.50	-.09	-.56	-1.45
	(.42)	(.98)	(1.35)	(1.16)

	Accountability Intervention: Norm-Referenced Test			
	White	Black	Hispanic	Asian
Math Teacher Expectations	2.93	2.25	2.63	3.81
	(.10)	(.14)	(.18	(.24)
NRT	-.20	-.34	-.69	1.40
	(.41)	(.84)	(.89)	(1.29)

	Accountability Intervention: Criterion-Referenced Test			
	White	Black	Hispanic	Asian
Math Teacher Expectations	2.94	2.27	2.63	3.81
	(.10)	(.14)	(.18)	(.24)
CRT	1.04	5.05	1.34	-.92
	(1.33)	(2.51)	(5.51)	(4.33)

	Accountability Intervention: Exit Exam			
	White	Black	Hispanic	Asian
Math Teacher Expectations	2.93	2.25	2.63	3.92
	(.10)	(.14)	(.18)	(.24)
Exit Exam	.31	3.96	1.62	1.03
	(.39)	(1.01)	(.87)	(1.44)

	Accountability Intervention: Testing Index			
	White	Black	Hispanic	Asian
Math Teacher Expectations	2.94	2.27	2.63	3.82
	(.10)	(.15)	(.18)	(.28)
Testing Index	.13	.47	.26	1.68
	(.25)	(.30)	(.44)	(.61)
Student Characteristics	✓	✓	✓	✓
Teacher Characteristics	✓	✓	✓	✓
School Characteristics[a]	✓	✓	✓	✓

Sources: State-Level Accountability Characteristics and the Education Longitudinal Study (2002 wave)
Notes: See Table 5.1 notes.
[a] Region is not included as an adjustment variable.

accountability intervention variables does not affect the sizes of the coefficients for math teacher expectations. In other words, math teacher expectations and accountability interventions are unrelated, but are associated with tenth grade math test scores.

Now consider the association between accountability interventions and math test scores. The coefficients for accountability interventions differ in both size and direction, thus indicating that the findings presented in Table 4.4 mask important differences in how students respond to accountability testing. For example, CRTs are associated with higher tenth grade math test scores for white, black, and Hispanic students, and lower test scores for Asian students. The sizes of the coefficients show that black students benefit the most from living in states that use CRTs. Exit exams are also associated with higher test scores for each race group. However, the coefficients show that the difference between states that do and do not use exit exams is largest for black students. The remaining panels provide additional evidence of how the association between accountability interventions and math test scores differs by student race.

Recall from the analysis presented in Table 4.5 that math teachers adjust their expectations in response to accountability interventions, such that they more strongly predict math test scores. Also recall the findings presented in Table 5.4, which provides evidence that math teachers tend to lower their expectations for the very students that education policy intends to help: low-performing black and Hispanic students. Given the evidence presented in Chapter 4 and the findings presented in this chapter, does student race determine whether teachers communicate accountability expectations to students?

Turn to Table 5.7, which shows results from models that include the interaction term between math teacher expectations and accountability interventions. Similar to the findings presented in Table 4.5, the coefficients for the interaction effects show that, in most cases, math teacher expectations more strongly predict tenth grade math test scores in states that use accountability interventions. The strength of the association, however, differs by student race. For example, compare main effects presented in Table 5.6 to those presented in Table 5.7. Math teacher expectations in states that use CRTs more strongly predict math test scores for white, black, and Hispanic students, but not Asian students. Similarly, math teacher expectations for Asian students in states that use PATs are less predictive of test scores. Together, the findings presented in Tables 5.6 and 5.7 show that math teacher expectations continue to be strong predictors of math test scores, but how much and for whom is a function of the type of accountability intervention used and student race.

Table 5.7: Coefficients for Math Teacher Expectations and Accountability Interventions, and the Interaction Between Math Teacher Expectations and Accountability Interventions from Regression Models Predicting 10th Grade Math Test, Separately for White, Black, Hispanic, and Asian Students

	Model 1			
	White	Black	Hispanic	Asian
Math Teacher Expectations (MTE)	2.93	2.26	2.63	3.81
	(.10)	(.14)	(.18)	(.24)

	Accountability Intervention: Early Implementer			
	White	Black	Hispanic	Asian
Math Teacher Expectations	2.93	2.33	2.66	3.65
	(.16)	(.32)	(.48)	(.47)
Early Implementer	-.02	1.00	.05	-.49
	(.49)	(.92)	(.92)	(1.54)
*MTE * Early Implementer*	.01	-.10	-.03	.19
	(.18)	(.35)	(.51)	(.52)

	Accountability Intervention: Performance Assessment Test			
	White	Black	Hispanic	Asian
Math Teacher Expectations	2.94	2.16	2.59	3.94
	(.12)	(.15)	(.18)	(.26)
PAT	.50	-.03	-.63	-1.57
	(.42)	(1.01)	(1.38)	(1.17)
*MTE * PAT*	-.05	.50	.40	-.78
	(.18)	(.39)	(.69)	(.46)

	Accountability Intervention: Norm-Referenced Test			
	White	Black	Hispanic	Asian
Math Teacher Expectations	2.91	2.23	2.43	3.54
	(.12)	(.17)	(.27)	(.30)
NRT	-.20	-.34	-.64	1.40
	(.41)	(.85)	(.88)	(1.28)
*MTE * NRT*	.03	.08	.34	.47
	(.16)	(.30)	(.36)	(.40)

	Accountability Intervention: Criterion-Referenced Test			
	White	Black	Hispanic	Asian
Math Teacher Expectations	2.40	2.00	-1.32	6.76
	(.40)	(.56)	(2.70)	(1.19)
CRT	1.11	4.65	1.65	-5.73
	(1.30)	(.2.93)	(3.70)	(4.34)
*MTE * CRT*	.56	.27	3.97	-2.96
	(.40)	(.58)	(.2.70)	(1.20)

	Accountability Intervention: Exit Exam			
	White	Black	Hispanic	Asian
Math Teacher Expectations	2.87	2.12	2.23	3.64
	(.15)	(.31)	(.30)	(.33)
Exit Exam	.31	3.94	1.67	1.01
	(.39)	(.99)	(.87)	(1.45)
*MTE * Exit Exam*	.12	.15	.55	.27
	(.17)	(.35)	(.38)	(.42)
	Accountability Intervention: Testing Index			
	White	Black	Hispanic	Asian
Math Teacher Expectations	2.94	2.28	2.63	3.85
	(.10)	(.14)	(.18)	(.24)
Testing Index	.13	.49)	.28	1.68
	(.25)	(.31)	(.44)	(.61)
*MTE * Testing Index*	.05	.10	.39	.22
	(.10)	(.13)	(.18)	(.19)
Student Characteristics	✓	✓	✓	✓
Teacher Characteristics	✓	✓	✓	✓
School Characteristics[a]	✓	✓	✓	✓

Sources: State-Level Accountability Characteristics and the Education Longitudinal Study (2002 wave)
Notes: See Table 5.1 notes.
[a] Region is not included as an adjustment variable.

SUMMARY

One of the main goals of accountability interventions is to pressure teachers to raise their expectations of traditionally low-performing groups, especially black and Hispanic students. Indeed, accountability systems are often justified with reference to the black-white test score gap, one of the most extensively studied group differences in education research and a perennial concern of education policymakers. This chapter examined how student race determines teacher expectations and whether teachers respond to accountability systems in ways that exacerbate achievement gaps between white, black, and Hispanic students.

I first investigated whether teacher expectations are based upon criteria that differ for students of different racial groups. The findings presented in Table 5.2 show that the determinants of teacher expectations differ according to a student's race. The largest disparities between race groups pertain to a student's academic past, which implies that teachers are more (or less) forgiving of students with troubled academic histories. For example, having been enrolled in remedial math or having been held back a grade have no impact on teacher expectations of Asian students, but are associated with lower teacher expectations for white, black, and Hispanic students. The findings also show that teacher and school characteristics influence math teacher expectations and that there are no differences in these effects across racial groups.

Next, I examined the association between accountability interventions and teacher expectations for each race group, and I find that CRTs decrease teacher expectations of black students and increase their expectations of Hispanic students. Table 5.5 offers a closer look at this association and shows that math teacher expectations of black and Hispanic students in states that use CRTs do not differ for high- and low-performing students. They also show that high-performing Hispanic students in non-CRT states are subject to even lower math teacher expectations than their low-performing same-race peers. These findings demonstrate that math teacher expectations of black and Hispanic students are influenced by stereotypes.

I then examined how race affects the association between math teacher expectations and math test scores. The findings presented in Table 5.5 show that the association between teacher expectations and math test scores is weakest for black and Hispanic students. A closer examination shows that teacher expectations are not associated with achievement gains between the tenth and twelfth grades for Hispanic students, whereas they are for black, white, and Asian students.

The final two pieces of analysis examined how accountability interventions affect student achievement. Table 5.6 shows that the association between accountability interventions and tenth grade math test scores differs by student race. I find that math test scores for black, Hispanic, and white students are higher in states that use CRTs and exit exams. I also find that testing frequency increases math test scores. Table 5.7 then shows that strength of the association between math teacher expectations and math test scores differs according to accountability status. I find that math teacher expectations of white, black, and Hispanic students more strongly predict math test scores in states that use CRTs, but not for Asian students.

Although not presented in this dissertation, I have also applied a counterfactual analysis to evaluate a hypothetical scenario: if teachers were to raise their expectations for students normally held to low-standards (i.e., students not expected to complete college versus those expected to complete college), would the within-race test score gap decrease? Findings from this analysis suggest that, if teachers were to raise their expectations, it would decrease within-race variation in test scores but would not sufficiently narrow the between-race gap in math test scores.

Considered together, the findings presented in this chapter show that student race is an important determinant of teacher expectations, and that using accountability interventions to pressure teachers to raise their expectations, especially for low-performing black students, fails to close the achievement gaps between white, black, and Hispanic students. State-level efforts to raise achievement are unsuccessful because teachers do not respond to them as the policymakers intended.

CHAPTER 6
CONCLUSION

I draw from the teacher expectations and accountability systems literature to construct a comprehensive model to examine the question: Does pressuring teachers to raise their expectations increase student achievement? I argue that prior research on accountability systems has failed to answer this question because it tends to be narrowly focused on the direct association between accountability interventions and student achievement. By contrast, I use data drawn from the Education Longitudinal Study matched to a unique state-level accountability systems dataset to examine the student-teacher relationship in a richer way than in prior research. The three empirical chapters offer detailed insight into the association between teacher expectations and student achievement. In particular, this dissertation offers a systematic assessment of how teachers respond to accountability interventions with regard to their expectations for students.

The first empirical chapter, Chapter 3, examines the student-teacher relationship exclusively. I show that math teacher expectations of students' future educational attainment are influenced by student characteristics, but also by teacher and school characteristics. I then show that teacher expectations are strong predictors of math test scores and student enrollment, net of the predictor variables. Chapter 4 then examines how teachers respond to accountability interventions and if they communicate accountability expectations to students. The findings show that teachers tend to decrease their expectations of students in response to accountability testing. Despite the negative influence on teacher expectations for students' future educational attainment, teacher expectations more strongly predict student achievement in states that use accountability tests. The results suggest that, in spite of their lower expectations, the association between teacher expectations and student achievement is stronger in these states *because* teachers adjust their expectations of students. In particular, the results suggest that teacher expectations are significantly lower for low-performing students. I argue this calibration effect is evidence that teachers adjust their expectations according to student performance. The third empirical chapter then addresses the claim that student race mediates how teachers respond to accountability expectations. The findings show that teachers hold black and Hispanic students to lower standards than white students, and that they selectively respond to accountability interventions. These lower expectations contribute to the achievement gaps between white, black, and Hispanic students.

Although these findings are largely consistent with prior research, they extend this body of research in important ways. Surprisingly, prior scholars and policymakers have failed to appreciate that pressuring teachers to raise their expectations has unanticipated and counterproductive consequences for the very students the policies are intended to help. I find that, rather than raising their expectations of students, teachers appear to use the information gathered from tests to lower their expectations or even to justify their already low expectations of students. Far from alleviating what former President Bush called the "soft bigotry of low expectations," accountability standards appear to produce a "hard bigotry of low expectations" or a "harsh reality of diminished expectations," insofar as students adopt their teachers' lower expectations as their own.

Do these findings support Merton's (1948) idea of a self-fulfilling prophecy? Assuming that students have the ability to learn and teachers' low expectations are inaccurate, then yes, the findings presented in this dissertation are evidence of a self-fulfilling prophecy. However, it is an insidious self-fulfilling prophecy because it is in the opposite direction than policymakers intend. Low teacher expectations are particularly harmful for black and Hispanic students because they hold high- and low-performing black and Hispanic students to equally low standards. For low-performing students, low teacher expectations may represent a reality that they cannot escape. Low expectations for high-performing students, however, are “harsh” in that compete with students’ hopeful aspirations. That is not to say that all calibration effects are all harmful. Teachers who calibrate their expectations to student performance may steer students toward more realistic or attainable goals. However, calibration effects are useful to the extent that are not biased and do not crush students’ own aspiration.

If teachers are not responding to accountability pressure in ways that policy intends, then what, in particular, are teachers responding to and what are they communicating to students? Prior research focuses primarily on the achievement gains associated with accountability system revisions or sanctions. This dissertation focuses on the type of information gathered about students as a proxy for accountability pressure, and how teachers respond to such information. For example, NRTs inform teachers how well their students perform relative to the national average. By contrast, CRTs inform teachers how well their students perform relative to state-defined achievement standards. Because no two states have the exact same accountability system and/or distribute test results in a similar manner (e.g., presentation of report card data), the impact of accountability expectations on student performance in one state cannot be extrapolated easily to performance in another state. To be sure, in an ideal world, we would obtain full information about how test scores are reported (e.g., state, school, teachers, community) and thereby locate the specific sanctions and/or rewards associated with the publication of the results. Although the data collected for this dissertation do not allow this level of identification of the measures, it does offer insight about how student performance is monitored (i.e., test type), which in turn informs teachers of the effectiveness of their own teaching. Given that policymakers are still fine-tuning accountability systems nationwide and that the research on accountability systems is exploring new means of examining the effect of accountability interventions on student performance, the analysis presented in this dissertation proposes new ways to examine the impact of accountability testing on students. In particular, it shows that teachers should not be ignored, given their keynote in translating state-level achievement expectations to expectations held for particular students.

Policy Implications

Whether used for assessments or accountability purposes, tests are here to stay. Since they were first implemented to monitor student performance in the 1970s, policymakers have slowly introduced incentives to increase student performance. Race achievement gaps have declined somewhat, but still persist. When he enacted NCLB, President Bush made it teachers' responsibility to raise student performance, especially that of black and Hispanic students. At the time, the discourse used among education policymakers took an accusatory tone. For example, "Teachers who believe that certain social groups are slower to learn – and react by lowering the bar for performance – rob those children of opportunities to grow intellectually and achieve their dreams" (U.S. Secretary of Education Paige March 2003). Teachers opposed, and even resisted, the NCLB demands because they pressured teachers to compromise their own curricula for a system defined by test-scores, but confirmed few resources to help teachers meet the everyday needs of their students.

Although pressuring teachers to increase their expectations of all students may not be an effective strategy to close achievement gaps, especially between white, black, and Hispanic students, that is not to say that tests are not useful tools. One of the arguments maintained by teachers is that policymakers are too far removed from schools to define realistic achievement expectations. Research following the enactment of NCLB and the findings presented in this dissertation suggest that introducing achievement expectations does not necessarily produce the anticipated results. For many low-performing schools, the achievement expectations are too demanding and the consequences for not meeting these expectations are too high, that they may resort to unorthodox behaviors, such as helping students cheat or falsifying test scores (Dewan 2010; Jacob and Levitt 2002) just to meet achievement standards. Rather than increasing the demands of teachers, policymakers need to consider the challenges that teachers face in the classroom and how tests can be used to identify the students and schools in need of additional help. Lightening some of the burden placed on teachers will make them feel less pressure to "teach to the test" or to go so far as to fabricate student results.

This is not to say that teachers should be permitted to hold students to different standards, or that teachers are not important. In their concluding comments about the growth of the race achievement gap between primary and secondary school, Phillips and colleagues (1998) state:

> This does not necessarily mean that *schools* are a major contributor to the black-white test score gap. Although blacks may attend worse schools than whites, may be treated differently from whites in the same schools, or may be less interested in school than initially similar whites, it is also possible that blacks' parenting practices, peer influences, summer learning opportunities, or beliefs about their academic ability could explain why they learn less between first and twelfth grades than initially similar whites.

The issue is not that the teachers are failing to do their jobs effectively as much as it is that teachers face too many challenges that accountability expectations ignore. Among many other things, teachers compete against limited school resources, low student motivation, and parents' expectations of students. Accordingly, improving student performance requires engaging all the parties involved: students, parents, educators, and policymakers. Teachers need to communicate to policymakers how the daily challenges they face in and out of the classroom prevent them from meeting achievement expectations. Improved communication between teachers and policymakers will not only help alleviate teachers' frustrations, but will enable policymakers to propose an accountability agenda that teachers support. In fact, the current administration is using NCLB as an example of how *not* to use tests and is approaching testing more cautiously. President Obama's policy has shifted away from sanctions and focuses more on working with teachers, allowing them the time to prepare to meet the needs of their students in accordance with achievement expectations. It also focuses more on rewards rather than sanctions, which relaxes some of the frustrations that teachers feel about meeting achievement demands (Abrams et al. 2003; Linn 1994; Taylor et al. 2003), and improves the motivation of both teachers and students. In theory, a system that rewards student performance encourages teachers to adopt hopeful expectations of students. By contrast, teachers in a system that sanctions poor performance are prone lower expectations because the costs of not meeting accountability expectations are too high.

Are hopeful expectations more effective than realistic expectations? At the level of social interaction, teachers communicate their expectations to students and students respond accordingly, and vice versa. For example, a study by Madon and colleagues (2001) shows that teacher expectations are self-fulfilling because their initial perceptions of students are based upon valid information. They argue that self-fulfilling prophecies exist simultaneously to student self-verification. In other words, teachers' initial perceptions influence students' own perceptions at the same time that students' initial perceptions influence teachers' perceptions of students. Thus, in order for hopeful expectations to be effective, teachers must be able to communicate them to students in a way such that the students *believe* that they are able to perform better. The increased emphasis on test scores not only undermines teachers' own beliefs about students, but also challenges their purpose in the classroom.

APPENDIX A

DETAILED DESCRIPTION OF THE EDUCATION LONGITUDINAL STUDY AND THE ANALYTICAL STRATEGY

Table A1 presents the distribution of the public school students according to the NCES status variable F1UNIV2B, which categorizes base-year wave students according to their 2004 enrollment status. I make an additional distinction between non-transfer and transfer students because ELS imputed twelfth grade math test scores for students who transferred schools. I used the information presented in Table A1 to construct an eight-category status variable: In school, in grade 12, non-transfer; In school, in grade 12, transfer; In school, out of grade 12, non-transfer; In school, out of grade 12, transfer; Homeschooled/out of scope; Early graduate; Dropout; and Nonrespondent/F1 status unknown. Table A2 presents the distribution by status of the teacher questionnaire component. Seventeen percent of the students have math teachers who did not complete a questionnaire, and of those whose teacher did complete a questionnaire, about eight percent have missing information for math teacher expectations.

Because the 2004 wave was conducted during the spring of 2004 when students were still enrolled in school, constructed a more specific enrollment status variable for the models that examine the likelihood of students graduating on time, being held back a grade, or dropping out of school. For this variable, I used the 2006 wave status variable F2F1GRD as well as the 2004 wave status variable F1UNIV2B to construct a graduation status variable. Table A3 presents the distribution of students according to their 2004 graduation status (used to construct an adjustment weight), and Table A4 presents a collapsed status variable that is used for analysis.

Table A1: Base-Year to First Follow-Up Enrollment Status (F1UNIV2B) by School Type

	Public		Catholic		Private	
	Non-transfer	Transfer	Non-transfer	Transfer	Non-transfer	Transfer
In school, in grade 12	8,808	803	1,652	140	1,063	135
In school, out of grade 12	126	90	3	9	6	7
Homeschooled/out of Scope	98		9		31	
Early Graduate	479		8		29	
Dropout	581		8		20	
Nonrespondent/F1 status unknown	976		75		80	
Total	11,068	893	1,755	149	1,229	142

Source: Education Longitudinal Study (2002 and 2004 waves)

Table A2: Base-Year to First Follow-Up Enrollment Status and Math Teacher Questionnaire Status (Public School Students Only)

	Completed Questionnaire: Non-Missing Teacher Expectations	Completed Questionnaire: Missing Teacher Expectations	Did Not Complete Questionnaire
In school, in grade 12, non-transfer	6,776	622	1,410
In school, in grade 12, transfer	569	76	158
In school, out of grade 12, non-transfer	77	13	36
In school, out of grade 12, transfer	59	6	25
Homeschooled/out of scope	61	13	24
Early graduate	352	53	74
Dropout	399	63	119
Nonrespondent/F1 status unknown	652	97	227
Total	8,945	943	2,073

Source: Education Longitudinal Study (2002 and 2004 waves)

Table A3: 2004 Graduation Status by School Type and Math Teacher Questionnaire Status (Public School Students Only)

	Completed Questionnaire and has Non-Missing Teacher Expectations	Completed Questionnaire but has Missing Teacher Expectations	Did Not Complete Questionnaire
Graduated on time	7,694	740	1,673
Held back a grade	166	28	78
Dropped out	399	63	119
Early graduate	352	53	74
Unknown/out of scope	334	59	129
Total	8,945	943	2,073

Source: Education Longitudinal Study (2004 and 2006 waves)

Table A4: Collapsed 2004 Graduation Status by Math Teacher Questionnaire Status (Public School Students Only)

	Completed Questionnaire and has Non-Missing Teacher Expectations	Completed Questionnaire but has Missing Teacher Expectations	Did Not Complete Questionnaire
Graduated early or on time	8,046	793	1,747
Held back a grade	166	28	78
Dropped out	399	63	119
Unknown/out of scope	334	59	129
Total	8,945	943	2,073

Source: Education Longitudinal Study (2004 and 2006 waves)

Missing Data and Sample Attrition

All item-specific missing data was imputed using Stata's best-subset regression command "impute" for all scaled variables and a similar program "implog" for all dichotomous variables. The "implog" program is identical to Stata's "impute" command except that it uses logistic regression rather than ordinary least squares regression. Given the sampling design, where students and teachers are nested within schools, I imputed missing student, teacher, and school data separately.

Students. Missing data for student characteristics was imputed for students who participated in both the base-year wave and first follow-up wave (N = 15,236). I imputed missing information for all variables using a list of composite variables – i.e., student's sex, race, mother's education (in years), father's education (in years), mother's occupational prestige (coded in 1989 GSS), father's occupational prestige (coded in 1989 GSS), family income (natural log), two-parent family, academic track, region, urbanicity, school type and school enrollment for the 2001-02 academic year. The student variables imputed include: remedial math, remedial English, ever held back, ever have behavior problem, disability (drawn from parent survey), and ninth grade GPA.[13] In addition to the composite variables, ninth grade GPA was imputed using the students' grade point average for the tenth, eleventh, and twelfth grades.

Schools. I collapsed the student-level to the school-level to impute for missing information at the school data (*N* = 751). The variables of interest include: region, urbanicity, school size, number of full-time teachers, and the percent minority, free lunch, remedial math, remedial English, academic track, general track, and vocational track. Except for region and urbanicity, which are composite variables provided by ELS, all school-level variables were drawn from the Common Core of Data (CCD) and are made available only in the restricted-use data.

All variables were imputed using a core set of variables: urbanicity, region, school type, tenth grade enrollment, and school mean of student family income (natural log). The variables imputed include school enrollment (CP02STEN), number full-time equivalent teachers (CP02FTE), percent minority (CP02MIN), and percent free lunch (CP02FLUN). In addition to the core set of variables, each variable was

[13] Both mother's and father's occupational prestige scores (in 1989 GSS scale) were recoded from their occupation composite variables. Recoding occupations into prestige scores results in missing information for parents who indicated that they had no job, were homemakers, or were in the military. There are 1,389 students with missing information for their mother's occupational prestige score and 769 with missing information for father's occupational prestige score. Prestige scores were imputed using the following composite variables: gender, race, region, urbanicity, school type, family structure (i.e., lived with two parents, mother only, father only, or other family structure), mother's education, father's education, and family income (natural log).

imputed using the raw variables to be imputed (i.e., school variables with missing data).

Teachers. Missing math teacher information was handled differently. Of the 15,236 students who participated in both the base-year and first follow-up waves, 12,925 have math teachers who completed a questionnaire (*N* = 9,888 public school students, 1,781 Catholic school students, and 1,256 private school students). However, there are 1,092 students (*N* = 943 public school students, 93 Catholic school students, and 56 private school students) who have missing information for the primary explanatory variable, math teacher expectations (see Table 2.2). Missing data was not imputed for math teacher expectations (*N* = 8,945), but was imputed for the teacher variables for teachers who completed a questionnaire (*N* = 12,925). Less than 1.5 percent of math teachers had missing information for the teacher characteristics: gender, race, age, highest degree and teaching experience. Information for each of these variables was imputed using the teacher characteristic variables, as well as region, urbanicity, school type, and school enrollment.

Adjustment Weights

I constructed three adjustment weights similar to those in Morgan and Todd (2008) to adjust for sample attrition. The first weight adjusts for restrictions imposed to define the base-year analysis sample. The second weight makes a further adjustment for participation in the 2004 wave, and the third weight adjusts for 2004 enrollment status. The adjustment weights were constructed in several steps, described below.

Base-Year Weight. The 2002 wave analysis sample is restricted to public school students who have valid non-missing information for the primary explanatory variable, teacher expectations, reducing the sample from 11,961 to 8,945 respondents.

There are two reasons why students may have missing information for this variable: (1) their math teacher did not complete a questionnaire ($N = 2{,}073$) or (2) their math teacher completed the questionnaire but has missing information for the primary explanatory variable ($N = 943$).

Restricting the sample to public school students, I first estimated a logit model that predicted the likelihood of respondents having non-missing information for the primary explanatory variable:

$$\text{Logit(D)} = X\hat{\phi}, \tag{2.1}$$

where includes the variables believed to be predictors of teachers completing a questionnaire – predictor variables include student-level variables (e.g., sex, race, socioeconomic status, family structure, academic track, and if the student was ever been held back, ever enrolled in remedial math, ever had a behavior problem, and if the student has a disability), teacher-level variables (e.g., sex, race, highest degree, and teaching experience), and school-level variables (e.g., region, urbanicity, school size, number of full-time teachers, as well as the percent of the student body that was minority, free lunch, college prep, general, vocational, and enrolled in remedial math).[14]

I then multiplied the NCES base-year poststratification weight (BYSTUWT) by the inverse probability of respondents being in the base-year wave analysis sample, $1/\hat{p}_1$, thus giving more weight to respondents who were not in the analysis sample.

---

[14] The likelihood of teachers completing a questionnaire is partially dependent upon their own characteristics. However, students whose teachers did not complete a questionnaire have missing information for all the teacher-level variables. I imputed all missing information for the teacher-level variables, and then estimated the probability of students having valid information for the primary explanatory variable. I estimated a preliminary model that did not include teacher-level information, and the mean and standard deviation are similar to the fully specified model. The comparison of the chi-squared statistics as well as the ranges suggests that teacher characteristics matter.

The specified model has a chi-squared statistics of 173.75 and 41 degrees of freedom. The extracted $\hat{p}_1$ has a mean and standard deviation of .75 and .12, respectively, with a minimum and maximum of .18 and .92.

2004 Follow-Up Weight. For models where the outcome variable is 12th grade math test scores or math gains (i.e., 12th – 10th grade math test score), I constructed a 2004 follow-up weight that adjusts for sample attrition. Using the base-year to 2004 follow-up wave status variable (see Table 2.2), I estimated a multinomial logit model predicting the likelihood of respondents being in school, in grade 12, non-transfer. I then multiplied the base-year weight adjustment weight by the inverse probability of being in the 2004 analysis sample, $1/\hat{p}_2$. The predictor variables include: sex, race, socioeconomic status, family structure, region, urbanicity, academic track, three 10th grade educational expectations variables (student, mother, and father), if the student was held back a grade, if the student had a behavior problem in the past, how often the student was late, absent, or skipped class, and if the student got into trouble, was suspended, put on probation, or transferred schools in the last year. The chi-squared test statistic for the fitted model is 3436.94 with 231 degrees of freedom. The extracted $\hat{p}_2$ has a mean and standard deviation of .74 and .17, and a minimum and maximum of .0072 and .96.

2004 Graduation Status Weight. I constructed a third adjustment weight used to model the association between teacher expectations and students' 2004 graduation status using the same steps as those used to construct the 2004 follow-up weight. The estimated a multinomial logit model has a chi-squared test statistic of 282.30 with 33 degrees of freedom. The extracted $\hat{p}_3$ has a mean of .96 with a standard deviation of .04, and a minimum and maximum of .38 and .99.

APPENDIX B

TEACHER EXPECTATIONS AND THE STUDENT ACHIEVEMENT: REVISITING A CLASS QUESTION

Table B1: 10th Grade Math Test Regressed on Math Teacher Expectations as well as the Adjustment Variables

	Model 1	Model 2	Model 3	Model 3b	Model 4
Math Teacher Expectations	2.77	2.82	2.83	2.87	2.12
	(.08)	(.08)	(.08)	(.07)	(.08)
Female	-3.40	-3.37	-3.39	-3.37	-3.98
	(.25)	(.24)	(.24)	(.24)	(.24)
Race					
Black	-8.52	-7.53	-6.35	-6.13	-5.35
	(.50)	(.51)	(.49)	(.44)	(.48)
Hispanic	-6.16	-5.26	-4.18	-4.42	-3.77
	(.44)	(.45)	(.49)	(.48)	(.47)
Asian	-.89	-.40	.08	-.77	-.04
	(.87)	(.85)	(.80)	(.63)	(.77)
Native American	-4.82	-4.18	-3.48	-4.46	-2.73
	(1.28)	(1.28)	(1.33)	(1.33)	(1.21)
Multiracial	-3.91	-3.34	-3.08	-2.66	-2.67
	(.71)	(.71)	(.69)	(.62)	(.67)
Family Background					
Mother’s education (in years)	.39	.37	.34	.29	.28
	(.08)	(.08)	(.07)	(.07)	(.07)
Father’s education (in years)	.28	.28	.25	.24	.17
	(.06)	(.06)	(.06)	(.06)	(.06)
SEI score of mother’s occupation in	.03	.02	.02	.01	.02
2002	(.01)	(.01)	(.01)	(.01)	(.01)
SEI score of father’s occupation in	.04	.04	.03	.03	.03
2002	(.01)	(.01)	(.01)	(.01)	(.01)
Family income (natural log)	.36	.36	.22	.24	.24
	(.20)	(.19)	(.19)	(.19)	(.18)
Two-parent family	.14	.06	-.02	-.13	-.45
	(.31)	(.30)	(.31)	(.32)	(.31)
Past History (as reported by parent)					
Learning disability	-6.06	-5.88	-6.05	-5.90	-5.71
	(.48)	(.48)	(.49)	(.46)	(.48)
Ever in remedial math	-3.19	-3.23	-3.29	-3.35	-3.12
	(.46)	(.45)	(.45)	(.43)	(.44)
Ever held back	-4.36	-4.29	-4.22	-4.03	-3.79
	(.42)	(.41)	(.40)	(.41)	(.40)
Ever have behavior problem	.47	.53	.68	1.03	1.58
	(.53)	(.53)	(.53)	(.55)	(.55)
Academic Track					
General	-2.52	-2.41	-2.20	-2.27	-1.52
	(.30)	(.30)	(.30)	(.27)	(.29)
Vocational	-2.53	-2.51	-2.15	-2.07	-1.58
	(.43)	(.43)	(.43)	(.42)	(.43)
Teacher Characteristics					
Female		-.15	-.34	-.40	-.45
		(.31)	(.31)	(.28)	(.29)
Black		-4.74	-3.71	-4.35	-3.01
		(.66)	(.70)	(.60)	(.67)
Hispanic		-2.42	-1.48	-1.18	-1.49
		(.82)	(.82)	(.93)	(.78)

	Model 1	Model 2	Model 3	Model 3b	Model 4
Asian		-2.15	-1.35	-1.20	-1.65
		(1.00)	(1.02)	(.97)	(.94)
Native American		-3.54	-3.16	-4.99	-3.47
		(1.01)	(1.20)	(1.76)	(1.25)
Multiracial		-3.88	-3.34	-2.66	-3.22
		(1.06)	(1.03)	(1.04)	(1.03)
Age (in years)		-.03	-.02	-.02	-.02
		(.02)	(.02)	(.02)	(.02)
Certified degree or higher		.53	.55	.96	.42
		(.35)	(.35)	(.32)	(.33)
Teaching Experience (in years)		.09	.09	.09	.08
		(.02)	(.02)	(.02)	(.02)
Urbanicity					
Urban			-.51		-.33
			(.47)		(.47)
Rural			.46		.52
			(.41)		(.41)
Region					
Northeast			.79		1.48
			(.54)		(.52)
South			1.27		1.35
			(.44)		(.43)
West			.99		.49
			(.60)		(.61)
School Composition					
School size			.00		.00
			(.00)		(.00)
Full-time teachers			-.00		-.00
			(.01)		(.01)
Percent minority			-.00		.01
			(.01)		(.01)
Percent free lunch			-.06		-.07
			(.02)		(.02)
Percent college prep			.01		.01
			(.01)		(.01)
Percent general			-.00		-.00
			(.01)		(.01)
Percent vocational			-.01		-.01
			(.01)		(.01)
Percent remedial math			-.03		-.01
			(.02)		(.02)
9th Grade GPA					3.94
					(.22)
Constant	49.01	49.02	47.76	48.41	47.59
R^2	.50	.51	.52	.58	.54

Source: Education Longitudinal Study (2002 wave)
Notes: See Table 3.4 notes.

Table B2: 12^{th} Grade Math Test Regressed on Math Teacher Expectations as well as the Adjustment Variables

	Model 1	Model 2	Model 3	Model 3b	Model 4
Math Teacher Expectations	3.28	3.35	3.34	3.42	2.52
	(.12)	(.11)	(.11)	(.10)	(.12)
Female	-3.89	-3.84	-3.87	-3.73	-4.64
	(.33)	(.33)	(.32)	(.30)	(.32)
Race					
Black	-7.95	-7.15	-6.08	-5.81	-4.66
	(.65)	(.69)	(.66)	(.57)	(.66)
Hispanic	-5.60	-4.74	-4.06	-3.68	-3.60
	(.59)	(.61)	(.65)	(.62)	(.51)
Asian	-.53	-.12	-.12	-.67	-.18
	(.98)	(.97)	(.91)	(.81)	(.91)
Native American	-4.85	-4.47	-3.66	-4.96	-2.31
	(1.95)	(1.93)	(2.13)	(1.86)	(2.09)
Multiracial	-3.04	-2.46	-2.36	-2.00	-1.84
	(.11)	(.85)	(.81)	(.72)	(.78)
Family Background					
Mother's education (in years)	.44	.44	.40	.29	.35
	(.11)	(.11)	(.10)	(.09)	(.10)
Father's education (in years)	.37	.34	.30	.31	.21
	(.08)	(.08)	(.08)	(.08)	(.08)
SEI score of mother's occupation in	.02	.01	.01	.00	.01
2002	(.02)	(.01)	(.01)	(.01)	(.01)
SEI score of father's occupation in	.06	.05	.05	.04	.04
2002	(.02)	(.02)	(.02)	(.02)	(.02)
Family income (natural log)	.34	.36	.22	.43	.26
	(.30)	(.29)	(.28)	(.25)	(.26)
Two-parent family	.71	.62	.48	-.01	.02
	(.42)	(.41)	(.40)	(.39)	(.41)
Past History					
Learning disability	-7.27	-7.07	-7.16	-7.02	-6.70
	(.62)	(.63)	(.61)	(.59)	(.64)
Ever in remedial math	-2.67	-2.67	-2.80	-3.37	-2.53
	(.54)	(.55)	(.54)	(.52)	(.56)
Ever held back	-4.67	-4.58	-4.44	-4.22	-4.04
	(.56)	(.56)	(.55)	(.54)	(.55)
Ever have behavior problem	-.01	.12	.19	.44	.90
	(.84)	(.85)	(.84)	(.73)	(.90)
Academic Track					
General	-1.99	-1.85	-1.60	-1.69	-.87
	(.37)	(.38)	(.37)	(.33)	(.37)
Vocational	-3.21	-3.12	-2.71	-2.56	-2.29
	(.65)	(.65)	(.62)	(.58)	(.62)
Teacher Characteristics					
Female		-.13	-.21	-.12	-.26
		(.38)	(.38)	(.35)	(.37)
Black		-3.55	-2.43	-3.71	-1.39
		(.94)	(.95)	(.81)	(.89)
Hispanic		-2.03	-1.29	-1.47	-1.67
		(1.00)	(1.08)	(1.27)	(1.02)

	Model 1	Model 2	Model 3	Model 3b	Model 4
Asian		-1.99	-1.31	-2.23	-1.69
		(1.29)	(1.23)	(1.24)	(1.11)
Native American		-2.81	-2.51	-6.45	-2.33
		(1.66)	(1.89)	(1.84)	(2.09)
Multiracial		-5.37	-4.77	-3.08	-4.79
		(1.61)	(1.62)	(1.50)	(1.77)
Age (in years)		-.02	-.01	-.00	-.01
		(.03)	(.03)	(.03)	(.03)
Certified degree or higher		.37	.18	.77	-.00
		(.43)	(.42)	(.38)	(.40)
Teaching Experience (in years)		.10	.10	.10	.09
		(.03)	(.03)	(.03)	(.03)
Urbanicity					
Urban			-.43		-.08
			(.58)		(.57)
Rural			.81		.89
			(.47)		(.46)
Region					
Northeast			.88		2.00
			(.74)		(.74)
South			.90		1.08
			(.55)		(.53)
West			.86		.28
			(.74)		(.75)
School Composition					
School size			.00		.00
			(.00)		(.00)
Full-time teachers			-.01		-.01
			(.02)		(.02)
Percent minority			-.00		.01
			(.02)		(.02)
Percent free lunch			-.07		-.08
			(.02)		(.02)
Percent college prep			.02		.02
			(.01)		(.01)
Percent general			-.00		-.00
			(.01)		(.01)
Percent vocational			-.01		-.01
			(.02)		(.01)
Percent remedial math			.01		.02
			(.02)		(.02)
9th Grade GPA					4.69
					(.29)
Constant	53.19	53.18	52.16	53.81	51.63
R^2	.53	.54	.55	.60	.58

Source: Education Longitudinal Study (2004 wave)
Notes: See Table 3.4 notes.

Table B3: Math Test Score Gains Regressed on Math Teacher Expectations as well as the Adjustment Variables

	Model 1	Model 2	Model 3	Model 3b	Model 4
Math Teacher Expectations	.49	.49	.48	.49	.32
	(.06)	(.06)	(.06)	(.06)	(.07)
Female	-.48	-.48	-.48	-.43	-.63
	(.20)	(.20)	(.19)	(.18)	(.19)
Race					
Black	.49	.37	.39	.44	.68
	(.28)	(.28)	(.33)	(.31)	(.34)
Hispanic	.74	.76	.52	.77	.61
	(.29)	(.31)	(.37)	(.37)	(.38)
Asian	.11	.05	-.37	-.29	-.38
	(.55)	(.55)	(.60)	(.61)	(.61)
Native American	.83	.88	.61	.60	.88
	(.89)	(.87)	(1.03)	(.94)	(1.01)
Multiracial	1.34	1.32	1.18	.90	1.29
	(.52)	(.51)	(.51)	(.45)	(.52)
Family Background					
Mother's education (in years)	.03	.03	.03	-.02	.02
	(.06)	(.06)	(.05)	(.05)	(.05)
Father's education (in years)	.12	.12	.11	.16	.09
	(.05)	(.05)	(.05)	(.05)	(.05)
SEI score of mother's occupation in	-.00	-.00	-.00	-.01	-.00
2002	(.01)	(.01)	(.01)	(.01)	(.01)
SEI score of father's occupation in	.00	.00	.00	-.00	-.00
2002	(.01)	(.01)	(.01)	(.01)	(.01)
Family income (natural log)	.05	.06	.02	.05	.03
	(.11)	(.11)	(.11)	(.12)	(.11)
Two-parent family	.52	.53	.50	.15	.41
	(.27)	(.27)	(.26)	(.25)	(.26)
Past History					
Learning disability	-.98	-.95	-.93	-.85	-.83
	(.35)	(.35)	(.35)	(.34)	(.35)
Ever in remedial math	-.10	-.06	-.10	-.38	-.05
	(.29)	(.29)	(.29)	(.30)	(.29)
Ever held back	-.47	-.45	-.33	-.34	-.25
	(.34)	(.35)	(.35)	(.33)	(.35)
Ever have behavior problem	-.40	-.43	-.48	-.67	-.34
	(.48)	(.48)	(.47)	(.46)	(.47)
Academic Track					
General	.50	.53	.52	.47	.67
	(.22)	(.22)	(.22)	(.21)	(.21)
Vocational	-.53	-.50	-.54	-.48	-.46
	(.40)	(.40)	(.40)	(.37)	(.40)
Teacher Characteristics					
Female		.01	.05	.11	.04
		(.22)	(.22)	(.23)	(.22)
Black		.52	.67	.83	.88
		(.33)	(.37)	(.49)	(.38)
Hispanic		.07	.04	-.46	-.03
		(.52)	(.55)	(.66)	(.52)

	Model 1	Model 2	Model 3	Model 3b	Model 4
Asian		.71	.54	.11	.46
		(.59)	(.58)	(.64)	(.56)
Native American		.95	.95	-.44	.98
		(.93)	(1.00)	(1.42)	(1.04)
Multiracial		.07	.28	1.49	.27
		(.72)	(.72)	(.85)	(.73)
Age (in years)		-.02	-.02	-.03	-.02
		(.01)	(.01)	(.02)	(.01)
Certified degree or higher		-.11	-.25	-.43	-.29
		(.21)	(.20)	(.23)	(.20)
Teaching Experience (in years)		.04	.05	.06	.04
		(.01)	(.01)	(.02)	(.01)
Urbanicity					
Urban			-.06		.02
			(.31)		(.31)
Rural			.13		.15
			(.28)		(.28)
Region					
Northeast			.13		.36
			(.35)		(.35)
South			-.30		-.26
			(.29)		(.29)
West			.22		.10
			(.35)		(.35)
School Composition					
School size			.00		.00
			(.00)		(.00)
Full-time teachers			-.00		-.00
			(.01)		(.01)
Percent minority			.01		.01
			(.01)		(.01)
Percent free lunch			-.02		-.02
			(.01)		(.01)
Percent college prep			.01		.01
			(.00)		(.00)
Percent general			.00		.00
			(.00)		(.00)
Percent vocational			.00		.00
			(.00)		(.00)
Percent remedial math			.03		.03
			(.02)		(.02)
9th Grade GPA					.95
					(.19)
Constant	4.13	4.10	4.25	4.57	4.14
R^2	.05	.05	.06	.19	.07

Source: Education Longitudinal Study (2004 wave)
Notes: See Table 3.4 notes.

Table B4: Coefficients for Math Teacher Expectations from Logistic Regression Models Predicting the Likelihood of Graduating on Time, Being Held Back, and Dropping Out of School by Spring 2004

	Outcome Variable: Graduate On Time				
	Unadjusted Model	Model 1	Model 2	Model 3	Model 4
Math Teacher Expectations	.49	.35	.36	.38	.26
	(.03)	(.04)	(.04)	(.04)	(.04)
	Outcome Variable: Held Back (Out of Grade 12)				
	Unadjusted Model	Model 1	Model 2	Model 3	Model 4
Math Teacher Expectations	-.34	-.23	-.23	-.24	-.14
	(.04)	(.06)	(.06)	(.06)	(.06)
	Outcome Variable: Dropout				
	Unadjusted Model	Model 1	Model 2	Model 3	Model 4
Math Teacher Expectations	-.52	-.38	-.40	-.41	-.30
	(.03)	(.04)	(.04)	(.04)	(.04)
Student Characteristics		✓	✓	✓	✓
Teacher Characteristics			✓	✓	✓
School Characteristics				✓	✓
9th Grade GPA					✓

Source: Education Longitudinal Study (2002 and 2006 waves)
Notes: The enrollment status variables are restricted to base-year wave respondents whose 2004 status was known, and therefore the sample is reduced to 8,611 students enrolled in 561 schools. For these variables, the data are weighted by the base-year adjustment weight multiplied by the inverse probability of remaining in the sample.

APPENDIX C

DO TEACHERS' RESPONSES TO ACCOUNTABILITY INTEREVENTIONS CONTRIBUTE TO ACHIEVEMENT GAPS BETWEEN WHITE, BLACK, AND HISPANIC STUDENTS?

The analysis presented in Chapter 5 examined the importance of student race as a determinant of teacher expectations, primarily focusing on four race groups: white, black, Hispanic, and Asian students. This appendix presents additional analyses for these race groups as well as results for Native American and multiracial students.

SUPPLEMENTARY ANALYSIS

Education policy such as NCLB was enacted in response to the prevalence of the achievement gap, namely the black-white test score gap. Table 5.3 presented findings where math teacher expectations were regressed on student race and accountability interventions. Table C1 restricts the analysis to black and white students only and re-estimates the models to investigate whether teachers' responses to black and white students are stronger than the findings in Table 5.3 suggest. The results presented in Table C1 show that teachers still expect black students to complete fewer years of schooling than their white peers. Teachers in states that use accountability interventions also fail raise their expectations of black students.

The findings presented in Table 5.2 show that the determinants of math teacher expectations differ by student race. By contrast, the analysis presented in Table 5.3 uses the full analytic sample, and includes interaction terms between race and accountability interventions. Table C2 shows the results when math teacher

Table C1: Math Teacher Expectations Regressed on Race, Accountability Interventions, and the Interaction Effect for White and Black Students

	Accountability Intervention: Early Implementer		
	Model 1	Model 2	Model 3
Black	-.76	-.84	-.46
	(.09)	(.15)	(.15)
Early Implementer	.04	.02	.13
	(.07)	(.07)	(.07)
*Black * Early Implementer*		.10	.24
		(.17)	(.16)
	Accountability Intervention: Performance Assessment Test		
	Model 1	Model 2	Model 3
Black	-.77	-.67	-.23
	(.09)	(.10)	(.10)
PAT	.04	.11	.11
	(.06)	(.07)	(.06)
*Black * PAT*		-.37	-.24
		(.17)	(.15)
	Accountability Intervention: Norm-Referenced Test		
	Model 1	Model 2	Model 3
Black	-.77	-.76	-.29
	(.09)	(.12)	(.11)
NRT	-.19	-.18	-.22
	(.06)	(.06)	(.06)
*Black * NRT*		-.04	-.01
		(.16)	(.15)
	Accountability Intervention: Criterion-Referenced Test		
	Model 1	Model 2	Model 3
Black	-.76	-.16	.60
	(.09)	(.43)	(.45)
CRT	-.23	-.19	.02
	(.14)	(.14)	(.13)
*Black * CRT*		-.60	-.89
		(.43)	(.44)

	Accountability Intervention: Exit Exam		
	Model 1	Model 2	Model 3
Black	-.77	-.78	-.15
	(.09)	(.14)	(.15)
Exit Exam	-.02	-.03	.07
	(.06)	(.06)	(.07)
*Black * Exit Exam*		.02	-.18
		(.16)	(.17)
	Accountability Intervention: Testing Index		
	Model 1	Model 2	Model 3
Black	-.77	-.77	-.29
	(.09)	(.09)	(.08)
Testing Index	-.09	-.10	-.07
	(.03)	(.04)	(.04)
*Black * Testing Index*		.01	-.02
		(.07)	(.06)
Student Characteristics	✓	✓	✓
Teacher Characteristics	✓	✓	✓
School Characteristics[a]	✓	✓	✓
9^{th} Grade GPA			✓

Sources: State-Level Accountability Characteristics and the Education Longitudinal Study (2002 wave)
Notes: The data are weighted by the NCES post-stratification weight (BYSTUWT) multiplied by the inverse probability of having non-missing information for the primary explanatory variable, math teacher expectations. N = 6,305 students enrolled in 536 schools.
[a] Region is not included as an adjustment variable.

Table C2: Accountability Intervention Coefficients from Regression Models Predicting Math Teacher Expectations, Separately for White, Black, Hispanic, and Asian Students

	Accountability Intervention: Early implementer			
	White	Black	Hispanic	Asian
Unadjusted Model	.13	.34	.13	-.33
	(.10)	(.16)	(.20)	(.33)
Model 1: Unadjusted Model + student and teacher characteristics	.09	.28	.06	-.13
	(.07)	(.16)	(.21)	(.29)
Model 2: Unadjusted Model + school characteristics[a]	-.07	.28	-.01	-.35
	(.09)	(.18)	(.21)	(.33)
Model 3: Model 1 + school characteristics[a]	.01	.22	-.04	-.32
	(.07)	(.17)	(.22)	(.30)
Model 4: Model 3 + prior academic performance	.10	.47	.16	-.09
	(.07)	(.16)	(.18)	(.23)

	Accountability Intervention: Performance Assessment Test			
	White	Black	Hispanic	Asian
Unadjusted Model	.14	-20	.25	-.34
	(.11)	(.17)	(.22)	(.30)
Model 1: Unadjusted Model + student and teacher characteristics	.09	-.23	.13	-.30
	(.07)	(.16)	(.23)	(.24)
Model 2: Unadjusted Model + school characteristics[a]	.14	-.22	.32	-.31
	(.09)	(.17)	(.22)	(.26)
Model 3: Model 1 + school characteristics[a]	.11	-.26	.22	-.20
	(.07)	(.15)	(.25)	(.24)
Model 4: Model 3 + prior academic performance	.11	-.18	.02	.07
	(.06)	(.14)	(.22)	(.21)

	Accountability Intervention: Norm-Referenced Test			
	White	Black	Hispanic	Asian
Unadjusted Model	-.28	-.25	-.18	.05
	(.09)	(.16)	(.17)	(.24)
Model 1: Unadjusted Model + student and teacher characteristics	-.24	-.23	-.16	.08
	(.06)	(.14)	(.17)	(.19)
Model 2: Unadjusted Model + school characteristics[a]	-.21	-.31	.18	.01
	(.08)	(.16)	(.25)	(.21)
Model 3: Model 1 + school characteristics[a]	-.18	-.24	.05	.01
	(.06)	(.14)	(.24)	(.19)
Model 4: Model 3 + prior academic performance	-.22	-.25	-.09	-.14
	(.06)	(.14)	(.22)	(.17)

	Accountability Intervention: Criterion-Referenced Test			
	White	Black	Hispanic	Asian
Unadjusted Model	.17	-1.55	.09	1.64
	(.21)	(.48)	(.58)	(1.23)
Model 1: Unadjusted Model + student and teacher characteristics	-.04	-.86	.27	2.19
	(.14)	(.43)	(.61)	(1.21)
Model 2: Unadjusted Model + school characteristics[a]	-.28	-1.08	.12	1.52
	(.18)	(.42)	(.61)	(.92)
Model 3: Model 1 + school characteristics[a]	-.20	-.68	.30	1.78
	(.13)	(.41)	(.60)	(1.00)
Model 4: Model 3 + prior academic performance	-.00	-.70	.21	.88
	(.12)	(.43)	(.53)	(.27)

	Accountability Intervention: Exit Exam			
	White	Black	Hispanic	Asian
Unadjusted Model	.13	-.06	.06	-.10
	(.09)	(.18)	(.20)	(.27)
Model 1: Unadjusted Model + student and teacher characteristics	.04	.00	.02	-.04
	(.06)	(.16)	(.20)	(.23)
Model 2: Unadjusted Model + school characteristics[a]	-.03	.05	-.01	-.15
	(.08)	(.18)	(.17)	(.25)
Model 3: Model 1 + school characteristics[a]	-.03	.07	-.02	-.31
	(.07)	(.16)	(.18)	(.23)
Model 4: Model 3 + prior academic performance	.07	-.03	.08	-.05
	(.07)	(.16)	(.17)	(.18)

	Accountability Intervention: Testing Index			
	White	Black	Hispanic	Asian
Unadjusted Model	-.06	-.09	-.03	.10
	(.05)	(.07)	(.09)	(.11)
Model 1: Unadjusted Model + student and teacher characteristics	-.10	-.07	-.05	.07
	(.04)	(.06)	(.09)	(.08)
Model 2: Unadjusted Model + school characteristics[a]	-.06	-.09	.10	.08
	(.05)	(.07)	(.10)	(.10)
Model 3: Model 1 + school characteristics[a]	-.10	-.08	.06	-.02
	(.04)	(.06)	(.10)	(.09)
Model 4: Model 3 + prior academic performance	-.07	-.08	.06	-.03
	(.04)	(.06)	(.08)	(.07)

Sources: State-Level Accountability Characteristics and the Education Longitudinal Study (2002 wave)
Notes: See Table 5.1 notes.
[a] Region is not included as an adjustment variable.

expectations are regressed on accountability interventions and the adjustment variables, separately for each race group. The results presented in Table C2 differ slightly from the results presented in Table 5.3, but follow a similar pattern.

NATIVE AMERICAN AND MULTIRACIAL STUDENTS

Table C3 presents the means and standard deviations for Native American and multiracial students (see Table 5.1 for the corresponding information for white, black, Hispanic, and Asian students). Similar to black and Hispanic students, Native American and multiracial students trail their white counterparts academically. However, both groups have similar socioeconomic status as white students. On average, Native American students attend smaller schools than the other groups.

Table C4 presents the coefficients when math teacher expectations are regressed on the student, teacher, and school predictor variables, and Table C5 presents the math teacher expectations coefficients for models for three outcome variables: tenth grade math test, twelfth grade math test, and math gains (i.e. 12^{th} – 10^{th} grade math test scores). Neither Table C4 nor C5 presents a model that adjusts for school fixed-effects because the sample sizes are too small and the students are concentrated in too few schools.

Table C3: Means and Standard Deviations of Primary Variables of Interest, Separately for Native American and Multiracial Students

	Native American		Multiracial	
	Mean	S.D.	Mean	S.D.
Achievement Test Scores				
IRT estimated number right (10th grade)	35.87	11.29	39.91	13.34
IRT estimated number right (12th grade)	38.84	14.47	44.94	14.95
Gain Score (12th – 10th grade)	4.29	6.27	5.43	6.57
Math Teacher Expectations (in years)	13.62	2.06	14.60	2.00
Female	.47		.48	
Family Background				
Mother's education (in years)	13.64	2.26	13.56	2.21
Father's education (in years)	13.54	2.54	13.69	2.65
SEI score of mother's occupation in 2002	47.57	13.18	44.95	12.82
SEI score of father's occupation in 2002	42.65	11.96	43.95	11.94
Family income (natural log)	10.47	1.07	10.60	.90
Two-parent family	.65		.67	
Academic Past				
Learning disability	.16		.15	
Ever in remedial math	.12		.13	
Ever held back	.20		.15	
Ever have behavior problem	.06		.10	
Academic Track				
General	.47		.44	
Vocational	.13		.11	
Teacher Characteristics				
Female	.42		.53	
Black	.02		.07	
Hispanic	.05		.03	
Asian	.04		.08	
Native American	.06		.02	
Multiracial	.00		.02	
Age (in years)	43.32	12.21	42.69	11.21
Certified degree or higher	.46		.56	
Teaching Experience (in years)	12.13	9.52	13.78	10.57
Urbanicity				
Urban	.17		.29	
Rural	.35		.19	
Region				
Northeast	.07		.21	
South	.21		.27	
West	.43		.31	
School Characteristics				
School enrollment	1083.62		1542.03	895.56
Full-time teachers	61.26	41.59	85.74	44.43
Percent minority	39.53	29.30	39.75	30.84
Percent free-lunch	27.41	18.43	21.12	14.87
Percent college prep	54.56	35.90	52.14	29.43
Percent general track	63.79	29.93	55.00	33.98
Percent vocational track	25.94	28.65	16.91	18.61
Percent remedial math	9.97	7.73	6.51	8.51
9th Grade GPA	2.36	.84	2.55	.78

Source: Education Longitudinal Study (2002 and 2004 waves)
Notes: N = 71 Native American and 438 multiracial students for all base-year wave variables. N = 50 Native American and 302 multiracial students for twelfth grade math test and math gains. See Table 3.1 notes.

Table C4: Math Teacher Expectations Regressed on Student, Teacher, and School Characteristics, Separately for Native American and Multiracial Students

	Model 1		Model 2	
	Native American	Multiracial	Native American	Multiracial
Female	.70	.40	.48	.36
	(.45)	(.19)	(.44)	(.19)
Family Background				
Mother's education (in years)	.15	.16	.22	.18
	(.09)	(.05)	(.13)	(.05)
Father's education (in years)	.02	.19	-.11	.17
	(.11)	(.04)	(.16)	(.04)
SEI score of mother's occupation in 2002	.03	.01	.03	.01
	(.02)	(.01)	(.02)	(.01)
SEI score of father's occupation in 2002	.02	-.00	.01	.00
	(.03)	(.01)	(.04)	(.01)
Family income (natural log)	.01	.07	-.03	.07
	(.21)	(.13)	(.20)	(.13)
Two-parent family	.34	.22	1.16	.23
	(.54)	(.23)	(.60)	(22)
Past History				
Learning disability	.50	-.95	.07	-1.07
	(.52)	(.27)	(.64)	(.25)
Ever in remedial math	.57	-.46	.39	-.39
	(.40)	(.29)	(.40)	(.30)
Ever held back	-1.16	-.74	-.29	-.76
	(.32)	(.31)	(.47)	(.30)
Ever have behavior problem	-1.11	-.65	-1.23	-.63
	(.50)	(.34)	(.76)	(.33)
Academic Track				
General	-1.31	-.48	-.78	-.50
	(.59)	(.19)	(.46)	(.20)
Vocational	-.42	-.75	-.46	-.74
	(.62)	(.31)	(.76)	(.34)
Teacher Characteristics				
Female			.90	-.10
			(.69)	(.21)
Black			.36	.57
			(.79)	(.39)
Hispanic			.82	1.09
			(1.03)	(.39)
Asian			-3.75	.26
			(.54)	(.23)
Native American			1.16	-.10
			(1.61)	(.54)
Multiracial			243.17	.81
			(88.36)	(.66)
Age (in years)			-.04	.01
			(.04)	(.02)
Certified degree or higher			-.14	-.08
			(49)	(.21)
Teaching Experience (in years)			.03	-.02
			(.04)	(.02)
Constant	13.89	14.93	12.81	14.96
R^2	.44	.32	.58	.35

Table C4: Continued

	Model 3		Model 4	
	Native American	Multiracial	Native American	Multiracial
Female	.31	.31	.27	.20
	(.55)	(.18)	(.50)	(.17)
Family Background				
Mother's education (in years)	.22	.15	.30	.08
	(.14)	(.05)	(.12)	(.05)
Father's education (in years)	-.08	.16	-.24	.09
	(.23)	(.04)	(.23)	(.04)
SEI score of mother's occupation in 2002	.02	.01	.01	.02
	(.02)	(.01)	(.02)	(.01)
SEI score of father's occupation in 2002	-.00	-.00	.01	-.00
	(.04)	(.01)	(.04)	(.01)
Family income (natural log)	-.04	-.01	-.03	-.02
	(.28)	(.14)	(.25)	(.14)
Two-parent family	1.30	-1.07	1.18	.22
	(.64)	(.26)	(.56)	(.21)
Past History				
Learning disability	-1.59	-1.07	-1.45	-.96
	(.74)	(.26)	(.63)	(.26)
Ever in remedial math	1.15	-.38	.90	-.11
	(.88)	(.29)	(1.13)	(.25)
Ever held back	.36	-.75	-.08	-.46
	(.65)	(.30)	(.76)	(.29)
Ever have behavior problem	-1.80	-.72	-.76	-.26
	(1.20)	(.28)	(1.39)	(.29)
Academic Track				
General	-51	-.43	-.61	-.11
	(.39)	(.21)	(.34)	(.19)
Vocational	.14	-1.16	-.53	-.64
	(.80)	(.31)	(.69)	(.28)
Teacher Characteristics				
Female	1.50	-.25	1.35	-.18
	(1.02)	(.21)	(.92)	(.18)
Black	2.37	.61	.60	.78
	(2.05)	(.43)	(1.68)	(.42)
Hispanic	2.24	1.01	1.54	1.57
	(1.91)	(.42)	(1.71)	(.43)
Asian	-4.37	.48	-4.83	.08
	(1.33)	(.40)	(1.54)	(.43)
Native American	2.71	-.30	3.47	.59
	(2.93)	(.47)	(2.72)	(.44)
Multiracial	312.24	.88	167.78	.70
	(.05)	(.49)	(94.73)	(.41)
Age (in years)	-.06	.02	-.04	.01
	(.05)	(.02)	(.05)	(.01)
Certified degree or higher	.55	-.22	.01	-.15
	(.47)	(.19)	(.56)	(.18)
Teaching Experience (in years)	.03	-.02	.02	-.02
	(.06)	(.02)	(.06)	(.01)

Table C4: Continued

	Model 3		Model 4	
	Native American	Multiracial	Native American	Multiracial
Urbanicity				
Urban	.13	.26	.48	.31
	(1.13)	(.23)	(.94)	(.23)
Rural	-.30	.10	-.29	.27
	(.79)	(.25)	(.78)	(.23)
Region				
Northeast	1.19	.38	.64	.22
	(1.17)	(.27)	(1.08)	(.25)
South	-.77	-.05	-.59	-.03
	(1.10)	(.28)	(1.01)	(.25)
West	.00	-.87	-.30	-1.02
	(.80)	(.27)	(.81)	(.26)
School Composition				
School size	-.00	.00	-.00	.00
	(.00)	(.00)	(.0)	(.00)
Full-time teachers	.06	-.02	.02	-.02
	(.04)	(.01)	(.04)	(.01)
Percent minority	-.04	.00	-.02	.01
	(.02)	(.01)	(.02)	(.01)
Percent free lunch	.03	-.02	.01	-.02
	(.04)	(.01)	(.04)	(.01)
Percent college prep	-.02	.00	-.00	.00
	(.01)	(.00)	(.01)	(.00)
Percent general	.01	.00	.00	-.00
	(.02)	(.00)	(.02)	(.00)
Percent vocational	.00	.01	-.01	.01
	(.02)	(.01)	(.02)	(.00)
Percent remedial math	-.09	-.01	-.04	-.00
	(.05)	(.01)	(.05)	(.01)
9th Grade GPA			.81	1.03
			(.45)	(.12)
Constant	12.06	15.15	12.79	14.94
R^2	.7199	.4027	.7652	.5090

Source: Education Longitudinal Study (2002 wave)
Notes: See Table C3 notes.

Table C5: Math Teacher Expectations Coefficients from Regression Models Predicting 10th and 12th Grade Math Test Scores and Math Test Gains, Separately for Native American and Multiracial Students

	Outcome Variable: 10th Grade Math Test	
	Native American	Multiracial
Unadjusted Model: Math teacher expectations	2.98	3.67
	(.70)	(.32)
Model 1: Unadjusted Model + student characteristics	2.16	2.87
	(.67)	(.38)
Model 2: Model 1 + teacher characteristics	2.60	3.02
	(.57)	(.38)
Model 3: Model 2 + school characteristics	2.58	3.06
	(.89)	(.38)
Model 4: Model 3 + prior academic performance	2.01	2.11
	(1.02)	(.40)
	Outcome Variable: 12th Grade Math Test	
	Native American	Multiracial
Unadjusted Model: Math teacher expectations	4.72	5.11
	(.91)	(.43)
Model 1: Unadjusted Model + student characteristics	3.20	3.76
	(1.04)	(.45)
Model 2: Model 1 + teacher characteristics	3.38	3.87
	(1.55)	(.47)
Model 3: Model 2 + school characteristics	4.09	4.01
	(2.41)	(.49)
Model 4: Model 3 + prior academic performance	4.22	3.22
	(2.80)	(.53)
	Outcome Variable: Math Gain (12th – 10th Grade Math Test)	
	Native American	Multiracial
Unadjusted Model: Math teacher expectations	.59	1.06
	(.49)	(.29)
Model 1: Unadjusted model + student characteristics	.35	.60
	(.67)	(.26)
Model 2: Model 1 + teacher characteristics	.16	.59
	(.74)	(.26)
Model 3: Model 2 + school characteristics	.30	.65
	(1.41)	(.28)
Model 4: Model 3 + prior academic performance	1.05	.71
	(1.51)	(.33)

Source: Education Longitudinal Study (2002 and 2004 waves)
Notes: The tenth grade math test score models are weighted by the NCES post-stratification weight (BYSTUWT) multiplied by the inverse probability of having non-missing information for the primary explanatory variable, math teacher expectations. The twelfth grade math test score and math gain score models are weighted by the base-year adjustment weight multiplied by the inverse probability of being enrolled in the twelfth grade and being in the same school as the base-year wave at the time of the 2004 wave. See Table C3 notes.

REFERENCES

Abrams, Lisa M., Joseph T. Pedulla, and George F. Madaus. 2003. Teachers' Opinions of Statewide Testing Programs." Theory into Practice 42:18-29.

Ainsworth-Darnell, James W. and Douglas B. Downey. 1998. "Assessing the Oppositional Culture Explanation for Racial/Ethnic Differences in School Performance." American Sociological Review 63:536-53.

Alexander, Karl L, Doris R. Entwisle, and Samuel D. Bedinger. 1994. "When Expectations Work: Race and Socioeconomic Differences in School Performance." Social Psychology Quarterly 57:283-99.

Amrein-Beardsley, Audrey and David C. Berliner. 2003. "Re-analysis of NAEP Math and Reading Scores in States with and without High-Stakes Tests: Response to Rosenshine." Education Policy Analysis Archives 11(25). Retrieved from http://epaa.asu.edu/epaa/v11n25/

Angrist, Joshua D. and Jonathan Guryan. May 2008. "Does Teacher Testing Raise Teacher Quality? Evidence from State Certification Requirements." *Economics of Education Review* 37:483-503.

Babad, Elisha Y. 1985. "Some Correlates of Teachers' Expectancy Bias." *American Educational Research Journal* 22:175-83.

Bank, Babara J., Ricky L. Slavings, and Bruce J. Biddle. 1990. "Effects of Peer, Faculty, and Parental Influences on Students' Persistence. *Sociology of Education* 63:208-25.

Bidwell, Charles E. and John D. Kasarda. 1975. "School District Organization and Student Achievement." *American Sociological Review* 40:55-70.

Braun, Carl. 1976. "Teacher Expectation: Sociopsychological Dynamics." *Review of Educational Research* 46:185-213.

Brenner, Aprile D., and Rashmita S. Mistry. 2007. "Congruence of Mother and Teacher Educational Expectations and Low-Income Youth's Academic Competence." *Journal of Educational Psychology* 99:140-53.

Brophy, Jere E. and Thomas L. Good. 1970. "Teachers' Communication of Differential Expectations of Children's Classroom Performance: Some Behavioral Data." *Journal of Educational Psychology* 61:365-74.

-------. 1974. *Teacher-Student Relationships: Causes and Consequences*. New York: Holt.

Carnoy, Martin and Susanna Loeb. 2002. "Does External Accountability Affect Student Outcomes? A Cross-State Analysis." *Educational Evaluation and Policy Analysis* 24:305-31.

Coleman, James S. 1961. *The Adolescent Society: The Social Life of the Teenager and its Impact on Education.* New York: Free Press.

Coleman, James S., et al. 1966. *Equality of Educational Opportunity*. Washington, DC: U.S. Government Printing Office.

Conant, James B. 1953. *Education and Liberty: The Role of Schools in Democracy*. Cambridge, MA: Harvard University Press.

Cooper, Harris M. 1979. "Pygmalion Grows Up: A Model for Teacher Expectation Communication and Performance Influence." Review of Educational Research 49:389-410.

Cooper, Harris M. and David Y. Tom. 1984. "Teacher Expectation Research: A Review with Implications for Classroom Instruction." *Elementary School Journal* 85:76-89.

Correll, Shelley J. 2001. "Gender and the Career Choice Process: The Role of Biased Self-Assessment." *The American Journal of Sociology* 106:1691-730.

Correll, Shelley J. and Cecilia L. Ridgeway. 2003. "Expectation States Theory." Pp. 29-51 in *The Handbook of Social Psychology*, edited by John Delamater. New York: Kluwer-Plenum Press.

Council of Chief State School Officers. 2002. "Key State Education Policies on PK-12 Education: 2002." Washington, DC.

Crosnoe, Robert, Shannon Cavanagh, and Glen H. Elder, Jr. 2003. "Adolescent Friendships as Academic Resources: The Intersection of Friendship, Race, an School Disadvantage." *Sociological Perspectives* 46:290-307.

Darling-Hammond, Linda. 1991. "The Implications of Testing Policy for Quality and Equality." *Phi Delta Kappan* 73:220-25.

-------. 2003. "Standards and Assessments: Where we are and what we need." *Teachers College Record*. Retrieved from http://www.tcredord.org/Content.asp?ContentID=11109.

Darling-Hammond, Linda, Barnett Berry, and Amy Thoreson. 2001. "Does Teacher Certification Matter? Evaluating the Evidence." *Educational Evaluation and Policy Analysis* 23:57-77.

Deere, Donald R. and Wayne Strayer. 2003. "Competitive Incentives: School Accountability and Student Outcomes in Texas. Working Paper.

Dewan, Shaila. 2010. "Georgia Schools Inquiry Finds Signs of Cheating." *New York Times*, February 11.

Diamond, John B., Antonia Randolph, and James P. Spillane. 2004. "Teachers' Expectations and Sense of Responsibility for Student Learning: The Importance of Race, Class, and Organizational Habitus." *Anthropology and Education Quarterl*y 35:75-98.

Downey, Douglas B. and Shana Pribesh. 2004. "When Race Matters: Teachers' Evaluations of Students' Classroom Behavior." *Sociology of Education* 77:267-82.

Dusek, Jerome B. 1975. "Do Teachers Bias Children's Learning?" *Review of Educational Research* 45:661-84.

Ehrenberg, Ronald G., Daniel D. Goldhaber, and Dominic J. Brewer. 1995. "Do Teachers' Race, Gender, and Ethnicity Matter? Evidence from the National Educational Longitudinal Study of 1988." *Industrial and Labor Relations Review* 48:547-61.

Farkas, George, Robert P. Grobe, Daniel Sheehan, and Yuan Shuan. 1990. "Cultural Resources and School Success: Gender, Ethnicity, and Poverty Groups within an Urban School District." *American Sociological Review* 55:127-42.

Ferguson, Ann A. 2000. *Bad Boys: Public Schools in the Making of Black Masculinity*. Ann Arbor, MI: University of Michigan Press.

Ferguson, Ronald F. 1998. "Teachers' Perceptions and Expectations and the Black-White Test Score Gap." Pp. 273-317 in *The Black-White Test Score Gap*, edited by Christopher Jencks and Meredith Phillips. Washington, DC: Brookings Institute.

-------. 2003. "Teachers' Perceptions and Expectations and the Black-White Test Score Gap." *Urban Education* 38:460-507.

Figlio, David N. 2005. "Names, Expectations, and the Black-White Test Score Gap." National Bureau of Economic Research, Inc.

Figlio, David N., and Lawrence Getzler. 2006. "Accountability, Ability, and Disability: Gaming the System?" *Advances in Applied Economics* 14:35-49.

Fuller, Bruce, Joseph Wright, Kathryn Gesicki, and Erin Kang. 2007. "Gauging Growth: How to Judge No Child Left Behind?" *Educational Researcher* 36:268-78.

Grissmer, David and Ann Flanagan. 1998. *Exploring Rapid Score Gains in Texas sand South Carolina*. Washington, DC: National Education Goals Panel.

Grodsky, Eric, John Robert Warren, and Erika Felts. 2008. "Testing and Social Stratification in American Education." *Annual Review of Sociology* 34:385-404.

Grodsky, Eric, John R. Warren, and Demetra Kalogrides. 2009. "State High School Exit Examinations and NAEP Long-Term Trends in Reading and Mathematics, 1971-2004." *Educational Policy* 23:589-614.

Hallinan, Maureen T. 2008. "Teacher Influences on Students' Attainment in School." *Sociology of Education* 81:271-83.

Hamilton, Laura. 2003. "Chapter 2: Assessment as a Policy Tool." *Review of Research in Education* 27:25-68.

Hamilton, Laura S., Brian M. Stetcher, Julia A. Marsh, Jennifer Sloan McCombs, Abby Robyn, Jennifer Lin Russell, Scott Naftel, and Heather Barney. 2007. "Standards-Based Accountability Under No Child Left Behind." MG-589-NSF. Santa Monica, CA: RAND Corporation.

Haney, Walter. 2000. "The Myth of the Texas Miracle in Education." *Education Policy Analysis Archives* 8(41). Retrieved from http://epaa.asu.edu/epaa/v8n41/

Hansen, Jan B. 1988. "The Relationship of Skills and Classroom Climate of Trained and Untrained Teachers of Gifted Students." Ph.D. Dissertation, Department of Educational Studies. Purdue University, West Lafayette, IN.

Hanushek, Eric A., John F. Kain, Jacob M. Markman, and Steven G. Rivkin. 2003. "Does Peer Ability Affect Student Achievement?" *Journal of Applied Economics* 18:527-44.

Hanushek, Eric A. and Margaret E. Raymond. 2003. "Improving Educational Quality: How Best to Evaluation our Schools?" Pp. 192-44 in *Education in the 21st Century: Meeting the Challenges of a Changing World*, edited by Y. Kodrzyci. Boston, MA: Federal Reserve Bank of Boston.

-------. 2004. "The Effect of School Accountability Systems on the Level and Distribution of Student Achievement." *Journal of the European Economic Association* 2:406-15.

-------. 2005. "Does School Accountability Lead to Improved Student Performance?" *Journal of Policy Analysis and Management* 24:297-327.

Hauser, Robert M. "Progress in Schooling: A Review." Pp. 271-318 in *Social Inequality*, edited by Kathryn M. Neckerman. Russell Sage Foundation Publications.

Harvey, Dale G. and Gerald T. Slatin. 1975. "The Relationships Between Child's SES and Teacher Expectations: A Middle-Class Bias Hypothesis." *Social Forces* 54:140-59.

Hedges, Larry V., and Amy Nowell. 1998. "Black-White Test Score Gap Convergence Since 1965." Pp. 149-181 in *The Black White Test Score Gap*, edited by Christopher Jencks and Meredith Phillips. Washington, DC: Brookings Institution Press.

Jacob, Brian A. 2001. "Getting Tough? The Impact of high School Graduate Exams." *Educational Evaluation and Policy Analysis* 23:99-122.

-------. 2005. "Accountability, Incentives, and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools." *Journal of Public Economics* 89:761-96.

Jacob, Brian A, and Steven D. Levitt. 2002. *Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating*. (NBER Working Paper 8968). Cambridge, MA: National Bureau of Economic Research.

Jones, M. G., Brett D. Jones, Belinda Hardin, Lisa Chapman, Tracie Yarbrough, and Marcia Davis. 1999. "The Impact of High-Stakes Testing on Teachers and Students in North Carolina." *Phi Delta Kappan* 81:199-203.

Jordan, Heather R., Robert Mendro, and Dash Weerasinghe. 1997. "Teacher Effect on Longitudinal Student Achievement: A Preliminary Report on Research On Teacher Effectiveness." *National Educational Longitudinal Study*. Washington, D.C.: U.S. Department of Education.

Jussim, Lee. 1989. "Teacher Expectations: Self-Fulfilling Prophecies, Perceptual Biases, and Accuracy." *Journal of Personality and Social Psychology* 57:469-80.

Jussim, Lee and Jacquelyne Eccles. 1992. "Teacher Expectations II: Construction and Reflection of Student Achievement." *Journal of Personality and Social Psychology* 63:947-61.

Jussim, Lee, Jacquelyne Eccles, and Stephanie Madon. 1996. "Social Perception, Social Stereotypes, and Teacher Expectations: Accuracy and the Question for the Powerful Self-Fulfilling Prophecy." *Advances in Experimental Social Psychology* 28:281-387.

Jussim, Lee and Kent D. Harber. 2005. "Teacher Expectations and Self-Fulfilling Prophecies: Knowns and Unknowns, Resolved and Unresolved Controversies." *Personality and Social Psychology Review* 9:131-55.

Kane, Thomas J. and Cecilia Elena Rouse. 1995. “Labor Market Returns to Two- and Four-Year College.” *The American Economic Review* 85:600-14.

Klein, Stephen P., Laura S. Hamilton, Daniel F. McCaffrey, and Brian T. Stetcher. 2000. “What Do Test Scores in Texas Tell Us?” *Education Policy Analysis Archives* 8:1-22.

Koretz, Daniel M. 2008a. *Measuring Up: What Educational Testing Really Tells Us.* Cambridge, MA: Harvard University Press.

-------. 2008b. “Further Steps Toward the Development of an Accountability-Oriented Science of Measurement.” Pp. 71-91 in *The Future of Test-Based Educational Accountability*, edited by Katherine E. Ryan and Lorrie A. Shepard. New York, NY: Routledge.

Koretz, Daniel M. and Sheila I. Barron. 1998. *The Validity of Gains on the Kentucky Instructional Results Information System (KIRIS)*. Santa Monica, CA: RAND.

Kozol, Jonathan. 1991. *Savage Inequalities: Children in America’s Schools*. New York: Crown.

Ladd, Helen F. 1999. “The Dallas School Accountability and Incentive Program: An Evaluation of its Impacts on Student Outcomes.” *Economics of Education Review* 18:1-16.

Lareau, Annette. 2003. *Unequal Childhoods: Class, Race, and Family Life*. Berkeley, CA: University of Berkeley Press.

Lee, Valerie and Susanna Loeb. 2000. “School Size in Chicago Elementary Schools: Effects on Teachers’ Attitudes and Student Achievement.” *American Educational Research Journal* 37:3-32.

Lee, Valerie E. and Julia B. Smith. 1995. “Effects of High School Restructuring and Size Gains in Achievement and Engagement for Early Secondary School Students.” *Sociology of Education* 68:241-70.

-------. 1996. "Collective Responsibility for Learning and its Effect on Gains in Achievement and Engagement for Early Secondary School Students." *Journal of Education* 104:103-47.

-------. 1997. "High School Size: Which Works Best, and for Whom?" *Educational Evaluation and Policy Analysis* 19:205-27.

Lillard, Dean R. and Philip P. DeCicca. 2001. "Higher Standards, More Dropouts? Evidence Within and Across Time." *Economics of Education Review* 20:459-73.

Linn, Robert L. 1994. "Performance Assessment: Policy Promises and Technical Measurement standards." *Educational Researcher* 23:4-14.

-------. 2000. "Assessments and Accountability." *Educational Researcher* 29:4-16.

-------. 2003. "Performance Standards: Utility for Different Uses of Assessments." *Education Policy Analysis Archives* 11(31). Retrieved March 2, 2011 (http://epaa.asu.edu/epaa/v11n31/).

-------. 2008. " Educational Accountability Systems." Pp. 3-24 in *The Future of Test-Based Educational Accountability*, edited by Katerine E. Ryan and Lorrie A. Shepard. New York, NY: Routledge.

Linn, Robert L, M. E. Graue, and Nancy M. Sanders. 1990. "Comparing State and District Test Results to National Norms: The validity of claims that 'everyone is above average.'" *Educational Measurement Issues and Practice* 9:5-14.

Madon, Stephanie, Alison Smith, Lee Jussim, Daniel W. Russell, Jacquelynne Eccles, Polly Pallumbo, and Michele Welkiewicz. 2001. "Am I as You See me or Do You See Me as I am? Self-Fulfilling Prophecies and Self-Verification." *Personality and Social Psychology Bulletin* 27:1214-24.

McDonnell, Lorraine M. 2004. *Politics, Persuasion, and Educational Testing.* Cambridge, MA: Harvard University Press.

Merton, Robert K. 1948. "The Self-Fulfilling Prophecy." *Antioch Review* 8:192-210.

Minnesota Population Center. 2007. "State High School Exit Examinations for Graduating Classes Since 1977." Retrieved July 12, 2007 (http://www.hsee.umn.edu/).

Morgan, Stephen L. 2005. *On the Edge of Commitment: Educational Attainment and Race in the United States*. Stanford: Stanford University Press.

Morgan, Stephen L. and Jennifer J. Todd. 2008. "A Diagnostic Routine for the Detection of Consequential Heterogeneity of Causal Effects." *Sociological Methodology* 38:231-81.

National Center for Education Statistics. 2010. "National Assessment of Educational Progress Data Explorer." Washington, DC. Retrieved January 11, 2011 (http://nces.ed.gov/nationsreportcard/naepdata/).

National Commission on Excellence in Education. 1983. "A Nation at Risk: The Imperative for Educational Reform." *The Elementary School Journal* 84:113-130.

Nunnally, Jum C. and Ira H. Bernstein. 1994. *Psychometric Theory*. New York: McGraw-Hill.

Office of Technology Assessment. 1992. "Testing in American Schools: Asking the Right Questions." Washington, DC: U.S. Government Printing Office.

Olson, Lynn. 2005. "Defying Predictions, State Trends Prove Mixed on Schools Making NCLB Targets." *Education Week* 25(2): 26-7.

Paige, Rod. 2003, March 12. "Remarks by Secretary Paige to the Commonwealth Club of California, Silicon Valley, CA. Retrieved July 16, 2007 (http://www.ed.gov/news/speeches/2003/03/03122003a.html)

Phillips, Meredith, James Crouse, and John Ralph. 1998. "Does the Black-White Test Score Gap Widen after Children Enter School?" Pp. 229-72 in *The Black-*

White Test Score Gap, edited by Christopher Jencks and Meredith Phillips. Washington, DC: Brookings Institute.

Porter, Andrew C., Robert E. Floden, Donald J. Freeman, William H. Schmidt, and John R. Schwille. 1988. "Content Determinants of Elementary School Mathematics." Pp. 96-113 in *Perspectives on Research on Effective Mathematics Teaching*, edited by D. A. Grows an T. J. Cooney. Hillside, NJ: Erlbaum.

Prime, Glenda M. and Rommel J. Miranda. 2006. "Urban Public High School Teachers' Beliefs about Science Learner Characteristics: Implications for Curriculum." *Urban Education* 41:506-32.

Raudenbush, Stephen W. 2004. *Schooling, Statistics, and Poverty: Can We Measure School Improvement?* Princeton, NJ: Educational Testing Service.

Resnick, Daniel P. 1982. "History of Educational Testing." Pp. 173-94 in *Ability Testing: Uses, Consequences, and Controversies, Part II*. Edited by A. K. Wigdor and W.R. Garner. Washington, DC: National Academy Press.

Riegle-Crumb, Catherine, George Farkas, and Chandra Muller. 2006. "The Role of Gender and Friendship in Advanced Course Taking." *Sociology of Education* 79:206-28.

Riehl, Carolyn and John W. Sipple. 1996. "Making the Most of Time and Talent: Secondary School Organization Climates, Teaching Task Environments, and Teacher Commitment." *American Educational Research Journal* 33:873-901.

Roderick, Melissa, and Mimi Engel. 2001. "The Grasshopper and the Ant: Motivational Responses of Low-Achieving Students to High-Stakes Testing." *Educational Evaluation and Policy Analysis* 23:197-227.

Roderick, Melissa, Brian A. Jacob, and Anthony S. Bryk. 2002. "The Impact of High-Stakes Testing in Chicago on Student Achievement in Promotional Gate Grades." *Educational Evaluation and Policy Analysis* 24:333-57.

Roderick, Melissa, Jenny Nagaoka, and Elaine Allensworth. 2005. "Is The Glass Half Full or Mostly Empty? Ending Social Promotion in Chicago." Pp. 223-59 in *Uses and Misuses of Data in Accountability Testing: 104th Yearbook of the National Society for the Study of Education*, edited by Joan L. Herman and Edward H. Haertel. Madlen, MA: Blackwell.

Rosenthal, Robert and Lenore Jacobson. 1968. *Pygmalion in the Classroom: Teacher Expectations and Student Intellectual Development*. New York: Holt.

Rubovits, Pamela and Martin Maehr. 1973. "Pygmalion Black and White." *Journal of Personality and Social Psychology* 25:210-18.

Schiller, Kathryn S. and Chandra Muller. 2003. "Raising the Bar and Equity? Effects of State High School Graduation Requirements and Accountability Policies on Students' Mathematics Course Taking." *Educational Evaluation and Policy Analysis* 25:299-318.

Sewell, William H., Archibald O. Haller, and Alejandro Portes. 1969. "The Educational and Early Occupational Attainment Process." *American Sociological Association* 34:82-92.

Sewell, William H., Archibald O. Haller, and George W. Ohlendorf. 1970. "The Educational and Early Occupational Attainment Process: Replication and Revision." *American Sociological Association* 35:1014-27.

Sewell, William H. and Robert M. Hauser. 1980. "The Wisconsin Longitudinal Study of Social and Psychological Factors in Aspirations and Achievements." Pp. 59-99 in *Research in Sociology of Education and Socializations: A Research Annual*, vol. 1, edited by A. C. Kerckhoff, Greenwich, CT: JAI.

Shepard, Lorrie A. and Katherine C. Dougherty. 1991. "Effects of High-stakes Testing on Instruction." In *The Effects of High Stakes Testing*, ed. R. L. Linn (Chair). Symposium presented at the annual meetings of the American Educational Research Association and the National Council on Measurement in Education, Chicago, IL: April.

Shepardson, Daniel P. and Edward L. Pizzini. 1992. "Gender Bias in Female Elementary Teachers' Perceptions of the Scientific Ability of Students." *Science Education* 76:147-53.

Shields, Patrick, M., Camille Esch, Andrea Lash, Christine Padilla, Katrina Woodworth, Katrina G. LaGuarda, and Nicolas Winter. 2004. "Evaluation of Title 1 Accountability Systems and School Improvement: First Year Findings." Washington, DC: US Department of Education.

Smith, Alison E., Lee Jussim, and Jacquelyne Eccles. 1999. "Do Self-Fulfilling Prophecy Accumulate, Dissipate, or Remain Stable Over Time?" *Journal of Personality and Social Psychology* 77:548-65.

Stetcher, Brian M., Sheila Barron, Tammi Chun, and Karen Ross. 2000. "The Effects of the Washington State Education Reform on Schools and Classrooms." CRESST Report 525. National Center for Research on Evaluations, Standards, and Student Testing (CRESST), Los Angeles, CA.

Steele, Claude. 1997. "A Threat in the Air: How Stereotypes Shape Intellectual Identify and Performance." *Personal and Social Psychology* 69:797-811.

Taylor, Grace, Lorrie Shepard, Freya Kinner, and Justin Rosenthal. 2003. "A Survey of Teachers' Perspectives on High-Stakes Testing in Colorado: What Gets Taught, What Gets Lost." CSE Technical Report 588. Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and

Student Testing (CRESST), Graduate School of Education & Information Studies, University of California, Los Angeles.

Tenenbaum, Harriet R., and Martin D. Ruck. 2007. "Are Teachers' Expectations Different for Racial Minority Than for European American Students? A Meta-Analysis." *Journal of Educational Psychology* 99:253-73.

Texas Education Agency. 2000. "Texas TAAS Passing Rates Hit Seven-year High; Four Out of Every Five Students Passes Exam." Press release, May 17.

Warren, John R., Krista N. Jenkins, and Rachael B. Kulick. 2006. "High School Exit Examinations and State-Level Completion and GED Rates." *Educational Evaluation and Policy Analysis* 28:131-52.

Wiley, Glenn and Arlene Eskilson. 1978. "Why Did You Learn in School Today? Teachers' Perceptions of Causality." *Sociology of Education* 51:261-69.

Winfield, Lorie F. 1990. "School Competency Testing Reforms and Student Achievement: Exploring a National Perspective." *Educational Evaluation and Policy Analysis* 12: 157-173.